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 finite volume methods for hyperbolic systems of balance laws

Irene Gómez Bueno

PhD Thesis

Advisors:

Carlos María Parés Madroñal, Manuel Jesús Castro Díaz

Universidad de Málaga

Noviembre 2022

UNIVERSIDAD DE MÁLAGA

AUTOR: Irene Gómez Bueno

https://orcid.org/0000-0001-9357-4609

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

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

Dª Irene Gómez Bueno, 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 finite volume methods for hyperbolic systems of balance laws,

realizada bajo la tutorización de Carlos María Madroñal y dirección de Manuel Jesús Castro Díaz y Carlos Parés Madroñal
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 19 de octubre de 2022

Fdo.: Irene Gómez Bueno Doctorando/a	Fdo.: Carlos Parés Madroñal Tutor/a
Fdo.: Carlos Parés Madroñal Director/es de tesis	Manuel Jesús Castro Díaz

D. Manuel Jesús Castro Díaz y D. Carlos María Parés Madroñal, Catedráticos del Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada de la Universidad de Málaga,

CERTIFICAN:

- Que Irene Gómez Bueno, con grado en Matemáticas y máster en Doble Título en Máster Universitario en ESO – Esp. Matemáticas / Máster Universitario Matemáticas, 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 finite volume methods for hyperbolic systems of balance laws.

- Que las publicaciones que avalan la tesis no han sido utilizadas en ninguna otra tesis doctoral ni lo serán en el futuro, puesto que todos los coautores de los trabajos, con la excepción de la doctoranda, son doctores.

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, 19 de octubre de 2022.

Dr. Manuel Jesús Castro Díaz

Dr. Carlos María Parés Madroñal

THESIS BY COMPILATION

This doctoral thesis has been is written in the format of collection of manuscripts (three journal papers and a book chapter), following the regulation of the "Escuela de Doctorado (ED-UMA)" of the University of Málaga.

The four manuscripts that support this thesis, playing the PhD student the double role of corresponding author and first author in all of them, are the following:

- **High-order well-balanced methods for systems of balance laws: a control-based approach.**
  I. Gómez-Bueno, M.J. Castro and C. Parés. Applied Mathematics and Computation 394 (2021): 125820. DOI: `https://doi.org/10.1016/j.amc.2020.125820`. Journal impact factor: 4.397 (7/267 in Applied Mathematics). CiteScore (Scopus): 2.67 (26/590 in Applied Mathematics, 10/167 in Computational Mathematics).

- **Well-Balanced Reconstruction Operator for Systems of Balance Laws: Numerical Implementation.**
  I. Gómez-Bueno, M.J. Castro and C. Parés. In: Muñoz-Ruiz, M.L., Parés, C., Russo, G. (eds) Recent Advances in Numerical Methods for Hyperbolic PDE Systems. SEMA SIMAI Springer Series, vol 28. Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-72850-2_3`

- **Collocation Methods for High-Order Well-Balanced Methods for Systems of Balance Laws.**
  I. Gómez-Bueno, M.J. Castro, C. Parés and G. Russo. Mathematics 9.15 (2021): 1799. DOI: `https://doi.org/10.3390/math9151799`. Journal impact factor: 2.592 (21/332 in Mathematics). CiteScore (Scopus): 2.9 (53/391 in General Mathematics).

- **Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws**
  I. Gómez-Bueno, S. Boscarino, M.J. Castro, C. Parés and G. Russo. Applied Numerical Mathematics, 184 (2023): 18-48 (2023). DOI: `https://doi.org/10.1016/j.apnum.2022.09.016`. Journal impact factor: 2.994 (31/267 in Applied Mathematics). CiteScore (Scopus): 3.7 (142/590 Applied Mathematics).

Agradecimientos

Aún me cuesta creer que esté escribiendo estas líneas. Hacer una tesis no es fácil pero, si te rodeas bien, hasta puedes disfrutar del camino. Y yo he tenido la suerte de rodearme de los mejores.

Siempre digo que nunca me había planteado hacer una tesis doctoral y, si tengo que nombrar a un "culpable" de haber llegado hasta aquí, ese es Carlos Parés. Cuando me tutorizó el Trabajo Fin de Grado despertó mi pasión por la Matemática Aplicada, y de qué manera... Muchas gracias por tu absoluta entrega, ayuda y dedicación durante estos años, no solo académicamente, sino también a nivel personal.

Igualmente, quiero dar las gracias a mi otro director, Manuel Castro, por su compromiso, apoyo e infinita paciencia durante estos años. Da igual lo ocupado que esté, siempre está ahí para ayudar a todos.

Cuando uno se rodea de los mejores, solo pueden salir cosas buenas. Gracias a todos los miembros del Grupo EDANYA: Mª Luz Muñoz, Tomás Morales, Jorge Macías, José Manuel González, José María Gallardo, Sergio Ortega, Marc de la Asunción, Carlos Sánchez, Cipriano Escalante y María López. Y, por supuesto, gracias a los maravillosos compañeros, muchos ya doctores, con los que he tenido la suerte de coincidir: Hugo Carrillo, Alejandro González, Ernesto Guerrero, Juan Carlos González, Ernesto Pimentel, Juan F. Rodríguez y Kleiton Schneider. Y, por supuesto, gracias a Celia por haber sido mi salvavidas en tantos momentos complicados. Amiga, es muy bonito compartir alegrías y agobios juntas.

Quisiera gradecer también a Giovanni Russo y Sebastiano Boscarino por los aportes en esta tesis y por hacer tan provechosa mi estancia en Catania. Y gracias también a Emanuele Macca por hacerme esos tres meses más llevaderos y por su generosidad.

Por último, tengo que dar las gracias a mi familia y, en especial, a mis padres. Mamá y papá, gracias por haberme apoyarme en todo y por vuestro amor incondicional. Recorrer este camino sabiendo que estábais detrás para sujetarme si hacía falta lo ha hecho mucho más fácil. Gracias también a todos mis amigos y amigas por haber esto estos años tan

especiales. Gracias a mis niñas Ana, Ele, Gloria y Paula, por ser refugio; a Fernan, por su apoyo incondicional y a Julia, por estar siempre a mi lado y cuidarme tan bien.

Gracias a todos y todas.

Contents

List of Figures

Resumen

Esta tesis, avalada por los trabajos [1], [2], [3] y [4] y presentada en la modalidad de Tesis por Compendio de Publicaciones, se sitúa en el contexto de la resolución numérica de sistemas hiperbólicos de leyes de equilibrio, una de las líneas de investigación más activas del grupo EDANYA (Ecuaciones Diferenciales, Análisis Numérico y Aplicaciones) de la Universidad de Málaga. Estos sistemas no lineales de ecuaciones en derivadas parciales de evolución, en los que están presentes un flujo y un término fuente, aparecen en numerosos modelos de la dinámica de fluidos: modelos de aguas someras, de fluidos multifásicos, de la dinámica de gases, de la magnetohidrodinámica, etc.

La mecánica de fluidos computacional representa en la actualidad uno de los recursos matemáticos más valiosos en la simulación de diversos fenómenos que ocurren en la naturaleza. Se ocupa de la simulación de la evolución de los fluidos por medio de la resolución numérica de problemas gobernados por sistemas de ecuaciones en derivadas parciales (EDPs) que describen su comportamiento. Sin embargo, el principal obstáculo radica en la imposibilidad de la resolución exacta de un gran número de estos problemas, lo que justifica la necesidad de utilizar métodos numéricos con los que poder entender, predecir e incluso controlar la evolución de los flujos de los fluidos estudiados. Estas herramientas matemáticas tienen importantes aplicaciones en un amplio abanico de áreas de estudio tales como la oceanografía, ingeniería hidráulica, biología, meteorología, climatología o aeronáutica, entre otras. Para llevar a cabo un correcto diseño de métodos numéricos para la resolución de estas EDPs será necesario, por tanto, un conocimiento profundo tanto de la propia naturaleza física de los fluidos a simular, como de las propiedades matemáticas de los sistemas que se estudian.

Una de las ecuaciones en derivadas parciales más generales para la descripción de la física que rige el movimiento de los fluidos son las conocidas ecuaciones de Navier-Stokes (ver [5]). Estas ecuaciones describen la conservación de la masa, la cantidad de movimiento y la energía, incluyendo una ecuación de estado que pone en relación la presión, la energía y la densidad. Estas famosas EDPs pueden simplificarse a través de ciertas hipótesis que conducen a otros modelos de menor complejidad como, por ejemplo, las ecuaciones de aguas someras o *shallow water*, también conocidas como ecuaciones de Saint-Venant, debido a que fueron deducidas en el caso unidimensional en 1843 por Jean Claude Barré de Saint-Venant. Este sistema de ecuaciones se ocupa de la descripción del movimiento de una capa de fluido con poco espesor. La deducción de estas ecuaciones en su versión

unidimensional es el resultado de promediar en la dirección vertical las de Navier-Stokes, suponiendo una serie de hipótesis:

- Se asume que el agua es homogénea e incompresible.
- Se supone que la presión tiene un comportamiento hidrostático.
- Se asume que la única fuerza interna que actúa en el fluido es la presión, despreciando así los efectos viscosos.
- Se supone que el fondo sobre el que evoluciona el fluido y la superficie libre pueden ser representados mediante la gráfica de una función que solo depende de una de las variables horizontales, x (y del tiempo t para el caso de la superficie libre).
- Se desprecian las variaciones verticales de la componentes horizontales de la velocidad, suponiendo que la velocidad del agua solo depende de x y de t.

Así, se llega a las ecuaciones de conservación de la masa y cantidad de movimiento para el modelo de aguas someras, que vienen dadas por:

$$\begin{cases} h_t + (hu)_x = 0, \\ (hu)_t + \left(hu^2 + \dfrac{gh^2}{2} \right)_x = ghH_x, \end{cases} \tag{1}$$

donde:

- $h = h(x,t) \geq 0$ es el grosor de la capa agua;
- $u = u(x,t)$ es la velocidad horizontal;
- g es la gravedad;
- $H(x)$ es la profundidad del fondo medido desde un nivel de referencia.

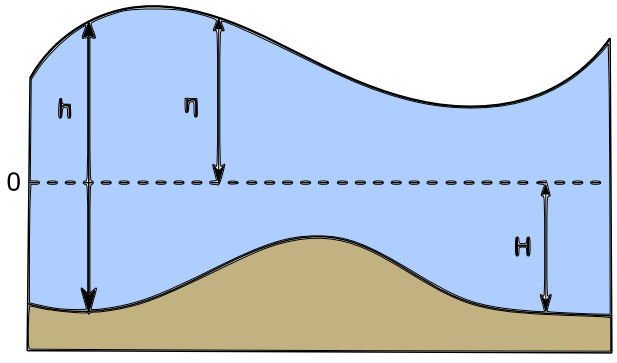


Figure 0.1: Boceto del modelo de agua someras.

Este sistema constituye un ejemplo de sistema hiperbólico de leyes de equilibrio, considerados en este trabajo. Otros ejemplos vienen dados por la ecuación de Burgers [6], las ecuaciones de Euler con gravedad [7], [8], el modelo de aguas someras con fricción [9] o las ecuaciones de Ripa [10], [11], entre otros.

Esta tesis doctoral está estructurada en dos bloques, constituyendo el primero de ellos el marco teórico en el que se encuadra la tesis, y el segundo la colección de los trabajos que la avalan. La tesis comienza, por tanto, ofreciendo una visión general ya existente en la literatura de los sistemas de leyes de equilibrio y su aproximación numérica en los capítulos 1 y 2. En concreto, la sección 1.1 se ocupa de la introducción de conceptos básicos relativos a estas ecuaciones: soluciones débiles y condiciones de Rankine-Hugoniot, entropía o problemas de Riemann y ondas simples, entre otros.

La resolución numérica de estos sistemas presenta diversas dificultades características. Una primera dificultad, que comparten con los sistemas de leyes de conservación (en los que no aparece el término fuente), viene de la aparición de discontinuidades en la solución débil, como son las ondas de choque. Los métodos numéricos han de ser capaces de capturar correctamente la aparición y evolución de estas discontinuidades. En esta tesis se considerarán métodos de tipo volúmenes finitos de alto orden basados en operadores de reconstrucción y en un flujo numérico robusto de primer orden: ver [12]. Dados los promedios de una función en las celdas de una malla, los operadores de reconstrucción proporcionan aproximaciones de alto orden de dicha función en cada una de las celdas. La aproximación en una celda se obtiene mediante técnicas interpolatorias usando los promedios en un conjunto de celdas vecinas, que se denominan *stencil*. A fin de capturar correctamente las discontinuidades y evitar oscilaciones en las proximidades de las mismas, estos operadores tienen que incorporar técnicas que permitan mantener controlada la variación total. Ejemplos de operadores clásicos de este tipo son los denominados ENO (essentially non-oscillatory), WENO (weighted essentially non-oscillatory) o CWENO (central weighted essentially non-oscillatory) (ver [13], [14]) basados en la interpolación polinómica.

La sección 1.2 se dedica al desarrollo de esquemas de alto orden para sistemas de leyes de equilibrio. Esta sección comienza con la discretización espacial de alto orden mediante el método de volúmenes finitos para sistemas de leyes de conservación. Para ello, se particiona el dominio del problema obteniendo una malla formada por un número finito de celdas computacionales cuyos extremos se conocen como interceldas. Posteriormente, se estudian pormenorizadamente los dos ingredientes clave de estos métodos: los flujos numéricos y los operadores de reconstrucción. En particular, se detalla la expresión de los flujos Lax-Friedrichs, su modificación conocida como flujo de Rusanov, también llamado flujo Lax-Friedrichs local, y el flujo HLL [15] (Harten, Lax y van Leer). En cuanto a los operadores de reconstrucción, se establece su definición y propiedades elementales y se presenta el diseño de las reconstrucciones MUSCL (monotone upstream centered scheme for conservation law), WENO y CWENO que han sido utilizadas en los trabajos que avalan esta tesis. Finalmente, se trata la extensión al diseño de métodos de volúmenes

finitos de alto orden para sistemas de leyes de equilibrio.

Cabe destacar que, una vez realizada la semidiscretización en espacio del sistema, se obtiene un esquema semidiscreto en tiempo que es, en realidad, un sistema de ecuaciones diferenciales ordinarias. El tipo de resolvedor en tiempo utilizado para la resolución de estas ecuaciones determinará el carácter del esquema: explicito, semi-implícito o implícito. En la sección 1.2.5 se detalla como realizar discretizaciones en tiempo explícitas, implícitas y semi-implícitas usando métodos de tipo Runge-Kutta (RK). Más concretamente:

- Para el caso de métodos explícitos se introducen los métodos de tipo Runge-Kutta TVD (total variation diminishing): ver [12]. Dado que en los tres primeros trabajos que avalan esta tesis se consideran las versiones de primer, segundo y tercer orden de estos métodos, las expresiones de estos integradores en tiempo se describen en dicha sección.

- Para el diseño de esquemas totalmente implícitos, se introducen los métodos Runge-Kutta diagonalmente implícitos (*diagonally implicit Runge-Kutta methods o DIRK*) y se muestra una particularización del método para el caso de segundo orden (ver [16]).

- Para el caso de esquemas semi-implícitos, se consideran métodos de tipo RK-IMEX (ver [17], [18], [19]), donde el tablero de Butcher para la parte implícita es de tipo DIRK. En determinadas situaciones una alternativa a los métodos RK-IMEX son los métodos propuestos en [20] y [21]. En este memoria hemos usado métodos de primer y segundo orden propuestos en [17].

En tres de los trabajos que apoyan la tesis ([1], [2], [3]) se consideran esquemas explícitos. Sin embargo, si el término hipérbolico y el término fuente de la ley de equilibrio son *stiff*, o solo alguno y/o parte de ellos, la resolución numérica mediante métodos explícitos implicaría pasos de tiempo muy pequeños que podrían hacer el problema inabordable desde el punto de vista práctico. En este caso, es necesario usar métodos implícitos o semi-implícitos, respectivamente. En la bibliografía podemos encontrar numerosas aplicaciones de esquemas implícitos o semi-implícitos como, por ejemplo, [22], [23], [24], [25], [26], [27], [28], [29] y [30]. El diseño de esquemas implícitos y semi-implícitos de alto orden que son, además, bien equilibrados, se ha abordado en el artículo [4], último de los trabajos que avalan esta tesis.

El capítulo 2 introduce una segunda dificultad, característica de los sistemas de leyes de equilibrio, que aparece cuando se desea simular ondas producidas por la perturbación de un estado de equilibrio: estos sistemas presentan soluciones estacionarias no triviales, que se corresponden con estados de equilibrio. Cuando estos equilibrios se perturban, se producen ondas que viajan a velocidad finita, permaneciendo inalterado el equilibrio en las zonas en las que aún no ha llegado. En muchos casos es importante que la solución numérica tenga esta misma propiedad. Para ello, es necesario que el método numérico preserve todas las soluciones estacionarias, o al menos las de cierta familia de especial

relevancia: se dice entonces que el método es bien equilibrado. El desarrollo de esquemas bien equilibrados es un frente muy activo de investigación actualmente y, en particular, en el grupo EDANYA: ver, por ejemplo, [31] y [32]. Son numerosos los trabajos que se encuentran en la literatura sobre el diseño de esquemas bien equilibrados: véase, por ejemplo, [33] y [34] para la ecuación de Burgers; [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60] y [61] para el sistema de aguas someras; [62] para el modelo de Ripa; [63], [64], [65], [66], [67], [68], [69], [70], [71][8], [72], [73], [74] y [75] para las ecuaciones de Euler compresibles con gravedad; o [76] y [77] para sistemas hipérbolicos generales.

En el contexto de los modelos de aguas someras, el desarrollo de métodos bien equilibrados se remonta al trabajo pionero de Bermúdez y Vázquez-Cendón [78] en el que se introducía la denominada propiedad-C de conservación: un método satisface dicha propiedad si preserva las soluciones estacionarias correspondientes al agua en reposo. Esta propiedad es crucial en aplicaciones como la simulación de tsunamis, en las que una onda de amplitud relativamente pequeña viaja sobre un estado de equilibrio en un dominio de gran escala.

El objetivo principal de esta tesis es el desarrollo de métodos numéricos generales de tipo volúmenes finitos bien equilibrados para sistemas de leyes de equilibrios escritos en la forma

$$U_t + f(U)_x = S(U)H_x, \tag{2}$$

donde la función H del término fuente se supone continua. Para ello, se consideran esquemas basados en el uso de operadores de reconstrucción usando la estrategia presentada en [32]. Dicha estrategia se basa en el uso de un operador de reconstrucción que es bien equilibrado en el siguiente sentido: cuando se aplica el operador a los promedios en las celdas de una solución estacionaria, las aproximaciones que proporciona en las celdas han de coincidir con dicha solución estacionaria. Dicho en otras palabras: las soluciones estacionarias que se desea preservar han de ser puntos fijos del operador de reconstrucción. Aunque esta no es, salvo casos muy particulares, una propiedad que posean los operadores de reconstrucción estándar, en [32] se propone una metodología para obtener un operador bien equilibrado a partir de uno que no lo es. Los pasos que se siguen para calcular la aproximación de una función en una celda conocidos sus promedios en la malla son los siguientes:

1. Se busca la solución estacionaria del problema cuyo promedio en la celda coincide con el de la función en la celda.

2. Se calculan en las celdas del stencil las diferencias entre los promedios de la función y los de la solución estacionaria hallada en el punto anterior. Una vez obtenidas estas diferencias, que llamaremos fluctuaciones, se aplica el operador de reconstrucción estándar que estemos considerando a las fluctuaciones.

3. Se toma como aproximación en la celda la suma de la solución estacionaria hallada en el primer punto y la reconstrucción hallada en el segundo punto.

Se demuestra que el operador que se obtiene así es bien equilibrado y del mismo orden que el seleccionado en el segundo paso siempre que las soluciones estacionarias sean suficientemente regulares. Nótese que el primer paso del procedimiento de reconstrucción implica la resolución, en cada celda, de problemas locales no lineales consistentes en encontrar una solución estacionaria en el stencil con promedio dado en las celdas. Esta estrategia ha sido aplicada con éxito para obtener esquemas que preservan todas las soluciones estacionarias para sistemas de leyes de equilibrio en los que se dispone de una expresión explícita o implícita de dichas soluciones. Tal es el caso de las ecuaciones de aguas someras (ver [79]) o del modelo de flujo sanguíneo en las venas (ver [80]). Los operadores de reconstrucción y los esquemas así obtenidos se conocen como exactamente bien equilibrados (*exactly well-balanced*).

El capítulo 2 se centra en el estudio de esquemas exactamente bien equilibrados obtenidos a través de operadores de reconstrucción que satisfacen esta propiedad. Para ello, se explica en detalle la metodología anterior que permite obtener un operador exactamente bien equilibrado a partir de uno de tipo estándar. Además, en la sección 2.2 se pone especial atención en las modificaciones que aparecen en el método como resultado del uso de fórmulas de cuadratura de alto orden para la aproximación tanto de los promedios como de la integral del término fuente. En particular, el uso de la fórmula del punto medio en esquemas de primer y segundo orden conduce a una expresión más simplificada de los mismos que se muestra en la sección 2.3. Este capítulo termina con el diseño de esquemas que son exactamente bien equilibrados para un conjunto de soluciones estacionarias en la sección 2.4.

Dado que el objetivo de esta tesis es el desarrollo de métodos numéricos generales de tipo volúmenes finitos bien equilibrados para cualquier sistema de leyes de equilibrio unidimensional, los trabajos que avalan este documento se ocupan del caso en el que no se dispone de una expresión analítica de las soluciones estacionarias del sistema o es muy costoso obtenerla. En este caso, la dificultad aparece en la resolución de los problemas locales no lineales del primer paso del proceso de reconstrucción que implican la búsqueda o evaluación de soluciones estacionarias con promedio dado. Esta dificultad se aborda en los tres primeros trabajos que avalan la tesis [1], [2], [3], en los que se propone la resolución numérica de estos problemas locales. Los operadores de reconstrucción y métodos así obtenidos se conocen como bien equilibrados (*well-balanced*), en lugar de exactamente bien equilibrados. No obstante, es inmediato comprobar que todo esquema exactamente bien equilibrado es también bien equilibrado.

En el caso de una dimensión de espacio, las soluciones estacionarias satisfacen un sistema de ecuaciones diferenciales ordinarias (ODE). Por tanto, el problema local a resolver en cada celda en el primer paso del proceso de reconstrucción consiste en encontrar la solución de un sistema de ecuaciones diferenciales con promedio dado en la celda. Se aproximará dicha solución usando resolvedores numéricos. De hecho, se verá que solo es necesario obtener aproximaciones en las interceldas y en los puntos de cuadratura de la fórmula seleccionada para calcular promedios e integrales en las celdas del stencil. Se

considerarán, por tanto, resolvedores de ecuaciones diferenciales ordinarias como parte de los ingredientes necesarios para abordar la resolución de los problemas locales. Fijémonos en que en este caso los operadores de reconstrucción serán de tipo discreto, proporcionando aproximaciones en los puntos de cuadratura y las interceldas de cada celda. Por tanto, se introduce una modificación del procedimiento de reconstrucción para obtener operadores de reconstrucción discretos que sean bien equilibrados.

A diferencia de los métodos exactamente bien equilibrados, los introducidos en los trabajos de esta tesis no preservan los promedios de las soluciones estacionarias exactas, sino aproximaciones obtenidas a través del uso de resolvedores numéricos de ODEs y de fórmulas de cuadratura. Resulta natural, por tanto, introducir el concepto de soluciones estacionarias discretas, definidas como secuencias de valores en las celdas que aproximan a los promedios de las soluciones estacionarias y que constituyen un equilibrio del sistema de ecuaciones diferenciales ordinarias correspondientes al método numérico semi-discretizado en espacio. Por tanto, se dice que un método numérico es bien equilibrado si tiene soluciones estacionarias discretas que aproximan a todas las soluciones estacionarias del problema o, al menos, a una familia relevante de ellas.

Los trabajos que avalan esta tesis presentan dos propuestas para la resolución numérica de los problemas locales no lineales de la primera etapa de reconstrucción y la obtención de soluciones estacionarias discretas: una de ellas basada en técnicas de control y la otra en la aplicación de métodos RK de colocación (ver [81]).

En primer lugar se propone interpretar los problemas locales a resolver en el primer paso como problemas de control escritos en forma funcional (ver [82]): hallar las condiciones iniciales (es decir, los valores en el extremo izquierdo de la celda) que hacen que el promedio en la celda de la solución de dicho sistema de ecuaciones diferenciales esté lo más próximo posible al valor prescrito. Es decir, la variable de control es el valor de la solución en el extremo izquierdo de la celda, las ecuaciones de estado son el sistema que satisfacen las soluciones estacionarias y el funcional de coste viene dado por la condición sobre el promedio (por ejemplo, la norma cuadrática de la diferencia entre el promedio de la solución del sistema y el valor dado).

La formulación de los problemas locales como problemas de control se encuentra en las dos primeras contribuciones introducidas y recogidas en el segundo bloque de esta tesis: el artículo publicado por I. Gómez-Bueno, M.J. Castro y C. Parés en 2021 en la revista *Applied Mathematics and Computation* [1], y el capítulo del libro de título "Recent Advances in Numerical Methods for Hyperbolic PDE Systems", redactado por los mismos autores y que forma parte de la serie SEMA SIMAI, publicado por la editorial *Springer* [2].

Es importante destacar que una dificultad específica de los problemas de control a resolver viene del hecho de que las ecuaciones de estado no están expresadas en forma normal, es decir, las derivadas de las incógnitas no aparecen despejadas. Además, en los estados sónicos (esto es, cuando alguno de los autovalores del jacobiano de la función flujo se anula) no se puede aplicar el Teorema de la Función Implícita para despejar la

derivada localmente. Esto hace que la preservación de soluciones estacionarias transónicas (es decir, que tienen al menos un estado sónico y cambian de régimen) sea especialmente difícil. Una estrategia para afrontar estos problemas y que permite trabajar correctamente con regímenes transónicos consiste en aplicar la reconstrucción estándar si se detecta un punto sónico en el stencil. Esta simple corrección ha sido utilizada con éxito en [1] y [2], permitiendo sortear esta dificultad y modificar el algoritmo de reconstrucción solo en los stencils en los que se detecten puntos sónicos.

Recordemos que los problemas de control están escritos en forma funcional, lo que nos permite obtener el gradiente del funcional a partir del problema adjunto. Uno de los principales objetivos de los trabajos [1] y [2] consiste en implementar estos problemas de control de forma eficiente mediante la aplicación del método de Newton o algoritmos de tipo gradiente con elección de paso adecuado. Como los problemas locales se resuelven en cada celda en cada paso de tiempo, la resolución eficiente de los problemas de control es crucial.

En estos trabajos se propone el uso de métodos explícitos de tipo RK para la resolución de los problemas, lo que introduce una nueva dificultad. Una vez aproximada la solución estacionaria en los puntos de cuadratura y las interceldas de la correspondiente celda, es necesario avanzar con el método RK seleccionado hacia la izquierda y hacia la derecha para obtener aproximaciones en los puntos de cuadraturas de las demás celdas del stencil. El carácter bien equilibrado se puede perder en este proceso si el resolvedor numérico de ODEs no satisface cierta propiedad de reversibilidad o simetría (ver [16]) de la que carecen los métodos explícitos. Este problema se resuelve mediante la reformulación del problema de control: se toma como variable de control el valor de la solución en el extremo izquierdo del stencil, con lo que solo es necesario avanzar hacia la derecha con el método numérico.

Las soluciones estacionarias discretas en este caso son las aproximaciones globales del sistema de ecuaciones diferenciales que satisfacen las soluciones estacionarias proporcionadas por el método RK explícito seleccionado. En [1] y [2] se considera como resolvedor el método de RK de cuarto orden.

En concreto, en [1] los problemas de control se resuelven mediante el método de Newton aplicando un algoritmo que se detalla en dicho artículo. Esta estrategia ha sido aplicada con éxito a un número importante de leyes de equilibrio, comenzando con tests de menor dificultad e incrementando la complejidad de los problemas. En particular, se han considerado:

- La ecuación de Burgers con dos tipos de término fuente no lineal, que constituye un caso particular de (2) para $N = 1$ correspondiente a

$$f(U) = \frac{U^2}{2}, \quad H(x) = x,$$

tomando

$$S(U) = U^2 \quad \text{o} \quad S(U) = \text{sen}(U).$$

En el primero de ellos es posible obtener una expresión analítica de las soluciones estacionarias que permite establecer una comparativa entre los métodos exactamente bien equilibrados desarrollados en [32] y los métodos bien equilibrados. En el segundo de ellos, sin embargo, no es posible obtener una expresión explícita o implícita de las soluciones estacionarias, por lo que se diseñan esquemas bien equilibrados que usan las técnicas de control introducidas.

- Las ecuaciones de Burgers acopladas con término fuente no lineal, que se corresponde con (2) para las elecciones $N = 2$,

$$U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad f(U) = \begin{pmatrix} \frac{u_1^2}{2} \\ \frac{u_2^2}{2} \end{pmatrix}, \quad S(U) = \begin{pmatrix} 2u_1^2 + u_1 u_2 \\ -u_1 u_2 + 3u_2^2 \end{pmatrix}, \quad H(x) = x.$$

- Las ecuaciones de aguas someras (1) o *shallow water*.

- Las ecuaciones de Euler con gravedad, caso particular de (2) con $N = 3$,

$$U = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad f(U) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E + p) \end{pmatrix}, \quad S(U) = \begin{pmatrix} 0 \\ -\rho \\ -\rho u \end{pmatrix}.$$

Especial relevancia tiene la aplicación en este último caso (ecuaciones de Euler con gravedad), ya que es la primera vez, hasta donde sabemos, que se ha diseñado una familia de métodos de alto orden que conservan soluciones estacionarias en movimiento para las ecuaciones de Euler con la gravedad.

Los experimentos numéricos realizados muestran que la modificación del operador de reconstrucción para que sea bien equilibrado incrementa, naturalmente, el coste computacional, especialmente para métodos de orden mayor o igual que tres. Sin embargo, este coste extra es menor que el que resultaría de considerar un método estándar que no sea bien equilibrado y refinar la malla para reducir los errores o incrementar el orden.

En el capítulo de libro [2] se establece una comparativa entre la aplicación del método de Newton y el uso de métodos de tipo gradiente para la resolución de los problemas de control. Como se ha dicho, dado que estos problemas se resuelven en cada celda en cada paso de tiempo, es imprescindible utilizar métodos que sean eficientes. Por ello, antes de llevar a cabo la comparativa entre ambas técnicas, se presentan dos propuestas de métodos de descenso: métodos de gradiente y de gradiente conjugado. Además, una elección de paso adecuada en estos métodos resulta una tarea fundamental. En este trabajo se discuten cinco propuestas para la elección de paso: paso fijo en cada iteración; dos versiones de paso variable en el que el valor del paso se multiplica o divide por un parámetro $\beta \in (0, 1)$ hasta que se satisfagan unas determinadas condiciones relacionadas con la minimización de la función objetivo; la regla de Armijo o las condiciones de Wolfe.

Se han considerado esquemas de tercer orden aplicando la fómula de cuadratura de Gauss de dos puntos como regla de aproximación de las integrales. Para comparar ambas estrategias se ha considerado un problema de control con ecuación escalar no lineal, y se han realizado distintas mediciones de los tiempos CPU y el número de iteraciones necesarias para resolver dicho problema un número elevado de veces (en concreto, se resuelve 10000 veces), tomando diferentes longitudes en la discretización espacial y distintas tolerancias de parada para los algoritmos. Los resultados muestran que el método de Newton es más eficiente que los métodos de gradiente, haciéndose esta diferencia más notable al refinar la discretización espacial y disminuir la tolerancia. Como cosecuencia, la obtención de métodos bien equilibrados para sistemas de leyes de equilibrio mediante técnicas de control se ha llevado a cabo a través de la aplicación del método de Newton. En concreto, se han considerado la ecuación de Burgers escalar con término fuente no lineal $S(U) = U^2$, y las ecuaciones de aguas someras. Además, para este último sistema, se ha considerado una variante del método de Newton en la que el estado adjunto solo se recalcula cada cierto número de iteraciones. Los resultados muestran que la mejor estrategia es resolver el problema adjunto solo una vez al comienzo del algoritmo de resolución del problema de control.

Como alternativa a las técnicas de control se propone también el desarrollo de esquemas bien equilibrados basados en el uso de métodos implícitos RK de colocación para resolver los problemas locales consistentes en un sistema de ecuaciones diferenciales con promedio dado. Este trabajo se recoge en el artículo publicado por I. Gómez-Bueno, M.J. Castro, C. Parés y G. Russo en 2021 en la revista *Mathematics* [3]. La ventaja del uso de estos métodos RK implícitos radica en su carácter simétrico o reversible, que evita la pérdida del carácter bien equilibrado de los métodos ya mencionada en el caso de las técnicas de control. Esta reversibilidad se deduce de la interpretación de los métodos como la búsqueda de un polinomio que satisface el sistema de ecuaciones diferenciales en los puntos de una fórmula de cuadratura. En este trabajo se consideran los métodos de colocación Gauss-Legendre cuyos nodos y pesos coinciden con los de las correspondientes fórmulas de cuadratura Gauss.

Esta estrategia ha sido aplicada con éxito en un número importante de leyes de equilibrio, comenzando con tests de menor dificultad e incrementando la complejidad de los problemas. En particular, se han considerado los siguientes problemas:

- La ecuación de Burgers con dos tipos de término fuente
$$S(U) = U^2 \quad \text{y} \quad S(U) = sen(U).$$
- Las ecuaciones de aguas someras (1) o *shallow water*.
- Las ecuaciones de aguas someras con fricción de Manning, dadas por
$$\begin{cases} h_t + q_x = 0, \\ q_t + \left(\dfrac{q^2}{h} + \dfrac{1}{2}gh^2\right)_x = ghH_x - \dfrac{kq|q|}{h^\eta}, \end{cases}$$

donde k es el coeficiente de Manning y η es un parámetro que tomamos igual a $\frac{7}{3}$.

- Las ecuaciones de Euler con gravedad, donde

$$U = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad f(U) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix}, \quad S(U) = \begin{pmatrix} 0 \\ -\rho \\ -\rho u \end{pmatrix}.$$

  La variable $\rho \geq 0$ representa la densidad, u es la velocidad, $p \geq 0$ la presión, E la energía total por unidad de volumen y $H(x)$ el potencial gravitatorio. La energía interna e está dada por $\rho e = E - \frac{1}{2}\rho u^2$. La presión se determina a partir de e mediante la ecuación de estado. En esta tesis, se considera la ecuación de estado para un gas ideal

$$p = (\gamma - 1)\rho e,$$

  donde $\gamma > 1$ es la constante adiabática.

De nuevo, los test muestran que aunque el diseño de esquemas que sean bien equilibrados es más costoso, son más efectivos que los métodos estándares cuando se aplican para simular la propagación de perturbaciones pequeñas de equilibrios o cuando se realizan simulaciones con tiempos largos cercanas a un equilibrio. Además, se han comparado los tiempos computacionales entre las distintas metodologías propuestas para obtener métodos bien equilibrados. Los resultados muestran que la estrategia basada en métodos de colocación es más eficiente que la desarrollada en los trabajos [1] y [2], en los que se utilizaba la formulación de los problemas locales como problemas de control.

Como ya se ha comentado, una importante dificultad de estos problemas se debe a que las ecuaciones diferenciales ordinarias que satisfacen las soluciones estacionarias no están, en general, expresadas en forma normal, es decir, las derivadas de las incógnitas no se encuentran despejadas. En [1] y [2] se proponía una estrategia para abordar estos problemas consistente en aplicar la reconstrucción estándar si se detecta un punto sónico en el stencil. En el artículo [3] se introduce una estrategia general que permite trabajar con problemas resonantes para sistemas leyes de equilibrio en una dimensión de la forma (2). Cuando el problema es resonante, se emplea una técnica en la que el valor de la solución estacionaria en un punto sónico se calcula mediante un proceso de paso al límite. En el caso de algunos sistemas como las aguas someras, este proceso se puede hacer analíticamente. En el artículo [3] se presentan dos experimentos para el modelo de aguas someras con fondo. En concreto, se ha considerado la simulación de una solución estacionaria transcrítica y su perturbación, que permiten comprobar la validez de la técnica introducida.

En los primeros tres trabajos de esta tesis, se aborda el diseño de esquemas bien equilibrados generales explícitos para sistemas de leyes de equilibrio generales. El último de los artículos que avalan esta tesis, publicado por I. Gómez-Bueno, S. Boscarino, M. J. Castro, C. Parés y G. Russo en 2022 en la revista *Applied Numerical Mathematics*

[4] se ocupa del desarrollo de métodos bien equilibrados de alto orden implícitos y semi-impícitos. Aunque, en principio, sería posible obtener esquemas de este tipo aplicando resolvedores numéricos implícitos o semi-implícitos para discretizar en tiempo los métodos semi-discretizados en espacio, la reconstrucción bien equilibrada llevaría a la aparición de complejos sistemas que serían inabordables. En particular, en cada paso de tiempo y en cada celda, el proceso de reconstrucción bien equilibrado implicaría la búsqueda de equilibrios locales relacionados con la incógnita en el siguiente paso de tiempo.

A fin de evitar la resolución de sistemas lineales muy complejos, en la estrategia que se sigue en [4] se escribe la aproximación numérica a calcular en cada celda en el tiempo t^{n+1} como la suma de la ya obtenida en el tiempo t^n y una fluctuación evaluada en el tiempo t^{n+1}. Esta fluctuación satisface un problema de Cauchy local con condición inicial nula en el tiempo t^n. El método avanza en tiempo mediante la resolución numérica de estos problemas de Cauchy en el intervalo $[t^n, t^{n+1}]$ usando métodos implícitos o semi-implícitos. La ventaja principal radica en que los operadores de reconstrucción que aparecen en las ecuaciones de las fluctuaciones pueden ser estándares: el carácter bien equilibrado se deriva del hecho de que, si la solución hallada en el tiempo t^n es una solución estacionaria discreta, entonces la solución del problema de Cauchy es idénticamente nula. Además, para reducir el coste de cálculo y la complejidad de los sistemas, se usan operadores simplificados en los que se usa solo la información que da la solución aproximada en el tiempo t^n para calcular los coeficientes y estimadores de regularidad necesarios.

Se plantea una estrategia general que permite el diseño de esquemas de alto orden, aunque numéricamente se han implementado métodos de primer y segundo orden. Los métodos propuestos se han probado numéricamente considerando varios experimentos numéricos y diferentes sistemas de leyes de equilibrio. En concreto, la sección numérica gira alrededor de tres sistemas: la ecuación lineal de transporte, la ecuación de Burgers y las ecuaciones de shallow water con topografía y fricción, obteniéndose resultados que confirman el carácter bien equilibrado de los métodos y su correcto funcionamiento para la simulación de estos problemas con algún término *stiff*.

En el segundo bloque de esta tesis se incluyen los cuatro manuscritos que la avalan, acompañados de un resumen y una breve discusión de los resultados obtenidos. Finalmente, se describen las conclusiones y trabajos futuros de la tesis.

Los trabajos de esta tesis se centran en el desarrollo de métodos numéricos de volúmenes finitos para sistemas generales de leyes de equilibrio en una dimensión escritos de la forma (2), donde H es una función continua. A continuación, realizaremos una breve descripción de algunas posibles líneas de investigación futuras que constituyen una continuación natural de los trabajos de esta tesis y que están incluidas en la sección dedicada a las conclusiones y trabajos futuros de la tesis. Uno de los trabajos futuros a considerar es la extensión de esta técnica a sistemas de leyes de equilibrio con término fuente no regular, es decir, en los que H tenga discontinuidades de salto. En una discontinuidad de H, se espera que la solución U también sea discontinua y el término fuente $S(U)H_x$ es un producto no conservativo al que dar sentido. Se pretende seguir la estrategia desarrollada en [32]:

siguiendo la teoría en [83], los productos no conservativos se interpretan como medidas de Borel cuya definición depende de la elección de una familia de caminos, que debe ser consistente con la física del problema. Se considerará la elección natural de estas familias de caminos para sistemas de leyes de equilibrio con término fuente singular descrita en [32].

Una segunda dificultad surge cuando se aproximan numéricamente sistemas hiperbólicos que dependen de un parámetro y que convergen a un sistema límite. Este sistema límite puede ser además de distinta naturaleza: puede ocurrir, por ejemplo, que cuando dicho parámetro tiende a cero el comportamiento en el límite sea parabólico, como ocurre en el caso de la ecuación hiperbólica del calor o en las ecuaciones de aguas someras cuando el inverso del coeficiente de fricción con el fondo tiende a cero. En esos casos, interesa tener métodos numéricos que preservan el comportamiento asintótico del modelo hiperbólico (*asymptotic-preserving*): es decir, métodos que, cuando el parámetro tiende a cero, son consistentes con el sistema asintótico y que son estables bajo los requisitos habituales para sistemas hiperbólicos. Cuando la velocidad de propagación de ondas tiende a infinito cuando el parámetro tiende a cero, es necesario recurrir a métodos implícitos o semi-implícitos para que cumplan el requisito de estabilidad: ver [84]. Una dificultad añadida en este caso es diseñar métodos numéricos que sean además bien equilibrados. En la actualidad, se están desarrollando esquemas *asymptotic-preserving* para sistemas que tienen por límite un sistema de leyes de equilibrio que además son bien equilibrados para el problema límite cuando el parámetro de relajación tiende a cero. Se dispone ya de prototipos que funcionan correctamente para problemas académicos y se está procediendo a su verificación en los modelos de interés.

Por otro lado, un importante reto aparece en la extensión esta técnica a sistemas multidimensionales. La dificultad aumenta considerablemente porque las soluciones estacionarias en este caso ya no satisfacen un sistema de ecuaciones diferenciales ordinarias, sino en derivadas parciales. Aunque los métodos RK no pueden ser extendidos fácilmente a la resolución de las ecuaciones en derivadas parciales que satisfacen las soluciones estacionarias en el caso multidimensional, en el caso particular de los métodos RK de colocación sí es posible extender su interpretación en términos de la búsqueda de polinomios que satisfacen la ecuación de forma exacta en los puntos de cuadratura y que minimicen una cierta distancia a los valores de las aproximaciones obtenidas en las celdas del stencil. Se considerarán también este tipo de técnicas de búsqueda de soluciones estacionarias próximas a los valores locales en el marco de los métodos Discontinuous-Galerkin. Se espera que estas estrategias puedan conducir al desarrollo de esquemas con propiedades de buen equilibrado para sistemas de leyes de equilbrio multidimensionales generales. El desarrollo de métodos de estas características es un problema abierto, por lo que cualquier avance en esta dirección tendría un alto impacto en la comunidad científica y un fuerte potencial de aplicaciones.

Abstract

This thesis, supported by the works [1], [2], [3] and [4] and presented in the modality of Thesis by Compilation, is framed within the context of the numerical resolution of hyperbolic systems of balance laws, one of the most active research lines of the EDANYA group (Differential Equations, Numerical Analysis and Applications) of the University of Málaga. These nonlinear systems of evolutionary partial differential equations, in which a flux and a source term are present, appear in numerous models of fluid dynamics: shallow water model, multiphase fluids, gas dynamics, magnetohydrodynamics, etc.

Computational Fluid Mechanics represents one of the most valuable mathematical resources in the simulation of various phenomena occurring in nature. For this purpose, the simulation of the evolution of fluids is developed by means of the numerical resolution of problems governed by systems of partial differential equations (PDEs) that describe their behavior. However, the main obstacle lies in the impossibility of exactly solving a large number of these problems, which justifies the need to use numerical methods which allow us to understand, predict and even control the evolution of the flow of the studied fluids. These mathematical tools have important applications in a wide range of study areas such as oceanography, hydraulic engineering, biology, meteorology, climatology or aeronautics, among others. In order to carry out a correct design of numerical methods for the resolution of these PDEs, it will be necessary, therefore, a deep knowledge of both the physical nature of the fluids to be simulated and the mathematical properties of the systems being studied.

One of the most general partial differential equations for describing the physics governing the motion of fluids is the well-known Navier-Stokes equations (see [5]). This system describes the conservation of mass, momentum and energy, including an equation of state that relates the pressure, energy and density. These famous PDEs can be simplified through certain hypotheses that lead to other models of less complexity, such as the shallow water equations, also known as the Saint-Venant equations, since they were derived in the one-dimensional case in 1843 by Jean Claude Barré de Saint-Venant. This system of equations deals with the description of the motion of a thin fluid layer. The derivation of these equations in their one-dimensional version is the result of averaging in the vertical direction the Navier-Stokes equations, assuming some hypotheses:

- Water is assumed to be homogeneous and incompressible.

- It is assumed that the pressure has a hydrostatic behavior.
- Horizontal viscous effects are neglected.
- It is assumed that the bottom on which the fluid evolves and the free surface can be represented by the graph of a function that only depends on one of the horizontal variables, x (and time t in the case of the free surface).
- The vertical variations of the horizontal components of the depth-averaged velocity are neglected, assuming that the water velocity only depends on x and t.

This leads to the equations of conservation of mass and momentum for the shallow water model, which are given by:

$$\begin{cases} h_t + (hu)_x = 0, \\ (hu)_t + \left(hu^2 + \dfrac{gh^2}{2} \right)_x = ghH_x, \end{cases} \tag{3}$$

where:

- $h = h(x,t) \geq 0$ is the thickness of the water layer;
- $u = u(x,t)$ is the horizontal velocity;
- g is the gravity;
- $H(x)$ is the bottom depth measured from a fixed reference level.

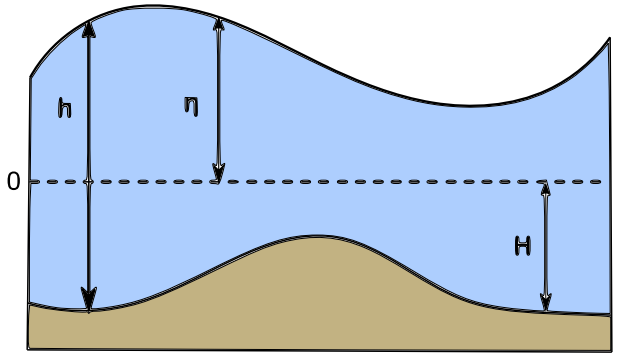


Figure 0.2: Sketch of the shallow water system.

This system constitutes an example of a hyperbolic system of balance laws, whose numerical resolution is the goal of this work. Other examples are given by the Burgers

equation [6], the Euler equations with gravity [7], [8], the shallow water model with friction [9] or the Ripa model [10], [11], among others.

This thesis is structured in two parts, the first one constituting the theoretical framework in which the thesis is based, whereas the second one includes the collection of the works that support it. The thesis begins, therefore, by providing an overview of general aspects already existing in the literature for systems of balance laws and their numerical approximation in Chapters 1 and 2. Specifically, Section 1.1 deals with the introduction of basic concepts related to these equations: weak solutions and Rankine-Hugoniot conditions, entropy conditions or Riemann problems and simple waves, among others.

The numerical resolution of these systems presents several difficulties. A first difficulty, which they share with systems of conservation laws (in which the source term does not appear), comes from the appearance of discontinuities in the weak solutions, such as shock waves. Numerical methods must be able to correctly capture the appearance and evolution of these discontinuities. In this thesis, high-order finite volume type methods based on reconstruction operators and a robust first-order numerical flux will be considered: see [12]. Given the averages of a function over the cells of a mesh, the reconstruction operators provide high-order approximations of that function in each of the cells. The approximation in a cell is obtained by interpolation techniques using the averages in a set of neighboring cells, which are called *stencil*. In order to correctly capture the discontinuities and avoid oscillations in the neighbourhood of discontinuities, these operators have to incorporate techniques to keep the total variation under control. Examples of classical operators of this type are the so-called ENO (essentially non-oscillatory), WENO (weighted essentially non-oscillatory) o CWENO (central weighted essentially non-oscillatory) (see [13], [14]) based on polynomial interpolation.

Section 1.2 is devoted to the development of high-order schemes for systems of balance laws. This section starts with the high-order spatial discretization using the finite volume method for systems of conservation laws. For this purpose, the problem domain is partitioned using a mesh composed by a finite number of computational cells whose extremes are known as intercells. Subsequently, the two key ingredients of these methods are studied in detail: the numerical fluxes and the reconstruction operators. In particular, the expression of the Lax-Friedrichs numerical flux, its modification known as the Rusanov or local Lax-Friedrichs flux, and the HLL [15] (Harten, Lax and van Leer) flux are detailed. Concerning the reconstruction operators, their definition and elementary properties are established and the design of the MUSCL (monotone upstream centered scheme for conservation law), WENO and CWENO reconstructions that have been used in the works that support this thesis is presented. Finally, the extension to the design of high-order finite volume methods for systems of balance laws is discussed.

It should be noted that, once the semi-discretization in space of the system has been carried out, a semi-discrete in time scheme is obtained, which is a system of ordinary differential equations (ODE). The type of numerical method used to solve this system will determine the character of the scheme: explicit, semi-implicit or implicit. Section 1.2.5

details how to perform explicit, implicit and semi-implicit in time discretizations using Runge-Kutta (RK) methods. More specifically:

- For the case of explicit methods, explicit Runge-Kutta total variation diminishing (TVD) methods are introduced: see [12]. Since first-, second- and third-order versions of these methods are considered in the works supporting this thesis, the expressions of these methods are described in that section.

- For the design of fully implicit schemes, diagonally implicit Runge-Kutta (DIRK) methods are introduced and a particular case of second-order method is shown (see [16]).

- In the case of semi-implicit schemes, RK IMEX methods are considered (see [17], [18], [19]), where the Butcher tableau for the implicit part is of DIRK type. Again, the particular case of a second-order method is given. In certain situations an alternative choice to is the one proposed in [20] and [21].

Explicit schemes are considered in three of the works supporting the thesis. However, if the hyperbolic term and the source term of the system of equilibrium laws are *stiff*, partially or in their entirety, their numerical resolution by explicit methods would involve very small time steps that may lead to an unapproachable problem from a practical point of view. In this case, implicit or semi-implicit methods will have to be adopted, respectively. Some relevant applications of these schemes can be found, for instance, in [22], [23], [24], [25], [26], [27], [28], [29] and [30]. The design of high-order implicit and semi-implicit schemes that are well-balanced has been developed in the last paper supporting this thesis [4].

Chapter 2 introduces a second difficulty, specific to systems of balance laws, which appears when one wishes to simulate the waves produced by the perturbation of an equilibrium state. When these steady states are perturbed, waves that travel at finite speed are produced, whereas the equilibrium remains unchanged in the areas where the perturbation has not yet arrived. In many cases it is important that the numerical solution has the same property. For this, it is necessary that the numerical method preserves all stationary solutions, or at least those of a certain family of special relevance: the method is then said to be well-balanced. The development of well-balanced schemes is a very active research line in Applied Mathematics and, in particular, in the EDANYA group: see, for example, [31] and [32]. Numerous papers can be found in the literature focused on the design of well-balanced schemes: see, for instance, [33] and [34] for the Burgers equation; [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60] and [61] for the shallow water system; [62] for the Ripa model; [63], [64], [65], [66], [67], [68], [69], [70], [71][8], [72], [73], [74] and [75] for the compressible Euler equations with gravity; or [76] and [77] for general hyperbolic systems.

In the context of shallow water equations, the development of well-balanced methods dates back to the pioneering work of Bermúdez and Vázquez-Cendón [78] in which the

so-called C-property of conservation was introduced: a method satisfies this property if it preserves the stationary solutions corresponding to water at rest. This property is crucial in applications such as tsunami simulations, where a relatively small amplitude wave travels over an equilibrium state in a large-scale domain.

The main objective of this thesis is the development of general finite volume well-balanced numerical methods for systems of balance laws written in the form

$$U_t + f(U)_x = S(U)H_x, \tag{4}$$

where the function H of the source term is assumed to be continuous. For this purpose, we consider schemes based on the use of reconstruction operators using the strategy presented in [32]. This strategy considers reconstruction operators that are well-balanced in the following sense: when the operator is applied to the cell averages of a stationary solution, the approximations that it provides in the cells coincide with that stationary solution. In other words, the stationary solutions to be preserved must be fixed points of the reconstruction operator.

Although this is not, except in very particular cases, a property fulfilled by standard reconstruction operators, in [32] a methodology was proposed to obtain a well-balanced operator from one that is not. The steps to follow to compute the approximation of a function in a cell given its averages in the mesh are:

1. Find the stationary solution of the problem whose average in the cell coincides with that of the function.

2. Compute the differences between the averages of the function and those of the stationary solution found in the step 1 in the stencil cells. Once these differences are obtained, which will be called fluctuations, the selected standard reconstruction operator is applied to the fluctuations.

3. The sum of the stationary solution found in the step 1 and the reconstruction found in the step 2 is taken as the approximation in the cell.

It is shown that the operator so obtained is well-balanced and of the same order as the standard one selected provided that the stationary solutions are smooth enough. Note that the first step of the reconstruction procedure involves solving, in each cell, local nonlinear problems consisting in finding a stationary solution with given average in a cell. This strategy has been successfully applied to obtain schemes that preserve all the stationary solutions for systems of balance laws in which an explicit or implicit expression of such solutions is available. This is the case for the shallow water equations (see [79]) or the blood flow model in veins (see [80]), among others. The reconstruction operators and the schemes so obtained are known as exactly well-balanced.

Chapter 2 focuses on the study of exactly well-balanced schemes obtained through reconstruction operators satisfying this property. For this purpose, the above methodology

that allows to obtain an exactly well-balanced operator from a standard one is explained in detail. In addition, in Section 2.2 special care is put to the modifications that appear in the method as a result of the use of quadrature formulas for the approximation of both the averages and the integral of the source term. In particular, the use of the midpoint rule in first- and second-order schemes leads to a simpler expression of the schemes shown in Section 2.3. This chapter ends with the design of schemes that are exactly well-balanced for a set of stationary solutions in Section 2.4.

Since the aim of this thesis is the development of general well-balanced finite volume numerical methods for any one-dimensional system of balance laws, the works supporting it deal with the case where the analytical expression of the stationary solutions of the system is not available or is very costly to compute. In this case, the difficulty appears in solving the local nonlinear problems in the first step of the reconstruction procedure involving the search for stationary solutions with given average and their evaluation. This difficulty is introduced in the first three works that support this thesis [1], [2], [3], and is approached by solving numerically the nonlinear problems of the reconstruction procedure. The reconstruction operators and methods thus obtained are called well-balanced, rather than exactly well-balanced. However, it is immediately shown that any exactly well-balanced scheme is also well-balanced.

In the case of one spatial dimension, the stationary solutions satisfy a system of ordinary differential equations. Therefore, the local problem to be solved at every cell in the first step of the reconstruction procedure consists of finding the solution of a system of differential equations with prescribed average in the cell. That solution will be approximated by ODE solvers. In fact, it will be shown that it is enough to obtain approximations at the intercells and at the quadrature points of the formula selected to compute the averages and integrals in the cells. ODE solvers will therefore be considered as part of the ingredients needed to address the solution of the local problems. Note that, in this case, the reconstruction operators will be of discrete type, providing approximations at the quadrature points and the intercells of each cell. Therefore, a modification of the reconstruction procedure is introduced to obtain discrete reconstruction operators that are well-balanced.

Unlike the exactly well-balanced methods, the ones introduced in this thesis do not preserve the averages of the exact stationary solutions, but approximations obtained through the use of numerical ODE solvers and quadrature formulas. It is natural, therefore, to introduce the concept of discrete stationary solutions, defined as sequences of values in the cells that approximate the exact averages of the stationary solutions and that constitute equilibria of the ODE system corresponding to the semi-discretized in space numerical method. Therefore, a numerical method is said to be well-balanced if it has discrete stationary solutions that approximate all the stationary solutions of the problem or at least a relevant family of them.

Two different strategies are introduced in the papers that support this thesis to solve the local problems arising in the first step of the well-balanced reconstruction procedure and

to compute the discrete stationary solutions: one of them is based on control techniques and the other one on the application of collocation RK methods (see [81]).

The first strategy is based on the interpretation of the local problems as control problems written in functional form (see [82]): find the initial condition (i.e. the value at the left intercell) such that the average in the cell of the solution of the Cauchy problem is as close as possible to the prescribed value. That is, the control variable is the value of the solution at the left intercell, the state equations are the ODE system satisfied by the stationary solutions, and the cost function is the distance between the average of the solution and the prescribed one in quadratic norm.

The formulation of local problems as control problems is studied in the first two contributions introduced and collected in the second part of this thesis: the paper published by I. Gómez-Bueno, M.J. Castro and C. Parés in 2021 in *Applied Mathematics and Computation* [1], and the chapter of the book entitled "Recent Advances in Numerical Methods for Hyperbolic PDE Systems", by the same authors published in the SEMA SIMAI series of *Springer* [2].

It is important to note that a specific difficulty of the control problems to be solved comes from the fact that the state equations are not expressed in normal form, i.e., the derivatives of the unknowns do not appear cleared. Furthermore, in sonic states (i.e., when any of the eigenvalues of the Jacobian of the flux function vanishes), the Implicit Function Theorem cannot be applied. This makes the preservation of transonic stationary solutions (i.e., having at least one sonic state and changing regime) particularly difficult. One strategy to cope with these problems is to apply the standard reconstruction if a sonic point is detected in the stencil. This simple correction has been successfully used in [1] and [2], allowing to circumvent this difficulty and to modify the reconstruction algorithm only in the stencils where sonic points are detected.

Since the local problems are solved in each cell at each time step, an efficient resolution of the control problems is crucial. Recall that the control problems are written in functional form, which allows us to obtain the gradient of the functional using the adjoint problem. One of the main objectives of the works [1] and [2] is to implement these control problems efficiently by applying Newton's method or gradient-type algorithms with an appropriate step choice.

In these works, the use of explicit RK-type methods is proposed for solving the state and adjoint ODE systems, which introduces a new difficulty. Once the stationary solution has been approximated at the quadrature points and the intercells of the cell where the average is prescribed, it is necessary to advance with the selected RK method to the left and to the right to obtain approximations at the quadrature points of the other cells of the stencil. The well-balanced character may be lost in this process if the numerical ODE solver does not satisfy a certain property of reversibility or symmetry (see [16]) that explicit methods lack. This problem is solved by reformulating the control problem: the value of the solution at the leftmost intercell of the stencil is taken as the control variable, so that it is only necessary to move to the right with the numerical method.

Discrete stationary solutions in this case are the global approximations of the ODE system satisfied by the stationary solutions provided by the selected explicit RK method. In [1] and [2] the fourth-order RK method is considered as ODE solver.

Specifically, in [1] the control problems are solved using Newton's method. This strategy has been successfully applied to a significant number of balance laws, starting with tests of lower difficulty and increasing the complexity of the problems. In particular, the following models have been considered:

- The Burgers equation with two types of nonlinear source terms, which constitutes a particular case of (4) for $N = 1$ corresponding to

$$f(U) = \frac{U^2}{2}, \quad H(x) = x,$$

taking

$$S(U) = U^2 \quad \text{or} \quad S(U) = \sin(U).$$

In the first case, it is possible to obtain an analytical expression of the stationary solutions what allows us to compare the exactly well-balanced methods developed in [32] and the well-balanced methods introduced here. In the latter, however, it is not possible to obtain an explicit or implicit expression of the stationary solutions, so well-balanced schemes using the introduced control techniques are applied.

- The coupled Burgers equations with a nonlinear source term, which correspond to (4) for the choices $N = 2$,

$$U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad f(U) = \begin{pmatrix} \frac{u_1^2}{2} \\ \frac{u_2^2}{2} \end{pmatrix}, \quad S(U) = \begin{pmatrix} 2u_1^2 + u_1 u_2 \\ -u_1 u_2 + 3u_2^2 \end{pmatrix}, \quad H(x) = x.$$

- The shallow water equations (3).

- The compressible Euler equations with gravity, which is a particular case of (4) with $N = 3$,

$$U = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad f(U) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E + p) \end{pmatrix}, \quad S(U) = \begin{pmatrix} 0 \\ -\rho \\ -\rho u \end{pmatrix},$$

where $\rho \geq 0$ is the density, u the velocity, $p \geq 0$ the pressure, E the total energy per unit volume and $H(x)$ the gravitational potential. The internal energy e is given by $\rho e = E - \frac{1}{2}\rho u^2$. Pressure is determined from e through the equation of state. In this thesis we assume an ideal gas law equation of state

$$p = (\gamma - 1)\rho e,$$

where $\gamma > 1$ is the adiabatic constant.

Of particular relevance is the application in the latter case (Euler equations with gravity), since it is the first time, to our knowledge, that a family of high-order methods that preserve moving stationary solutions for the Euler equations with gravity has been designed.

The numerical experiments show that the modification of the reconstruction operator to be well-balanced evidently increases the computational cost, especially for methods of order greater than or equal to three. However, this extra cost is less than that which would result from considering a standard method that is not well-balanced and refining the mesh to reduce the errors or increase the order when simulating the waves generated by a perturbation of a steady state.

In the book chapter [2] the efficiency of Newton's method and gradient-type methods for solving the control problems are compared. As mentioned, since these problems are solved in each cell at each time step, it is essential to use methods that are efficient. Therefore, before carrying out the comparison between the two techniques, two proposed descent methods are presented: gradient and conjugate gradient methods. Furthermore, an appropriate step choice in these methods is a fundamental task. Five proposals for the step choice are discussed in this work: fixed step in each iteration; two versions of variable step in which the step value is multiplied or divided by a parameter $\beta \in (0, 1)$ until certain conditions related to the minimization of the objective function are satisfied; Armijo's rule or the Wolfe's conditions.

Third-order schemes have been considered by applying the two-point Gauss quadrature formula to approximate the integrals. In order to compare both strategies, a control problem with a nonlinear scalar equation has been considered, checking the CPU times and the number of iterations needed to solve this problem a large number of times (specifically, it is solved 10000 times) taking different lengths in the spatial discretization and different stopping tolerances for the algorithms. The results show that Newton's method is more efficient than the gradient methods, with this difference becoming more noticeable as the spatial discretization is refined and the tolerance decreases. As a consequence, obtaining well-balanced methods for systems of balance laws by means of control techniques has been carried out through the application of Newton's method. Specifically, the scalar Burgers equation with the nonlinear source term $S(U) = U^2$, and the shallow water equations have been considered in this book chapter. In addition, for the latter system, a modification of the Newton's method has been considered in which the adjoint state is only recalculated every few iterations. The results show that the best strategy is to solve the adjoint problem only once at the beginning of the algorithm for solving the control problem.

As an alternative to the application of control techniques, we propose the development of well-balanced schemes based on the use of implicit collocation RK methods to solve local problems consisting of an ODE system with given average. This work is reported in the paper published by I. Gómez-Bueno, M.J. Castro, C. Parés and G. Russo in 2021 in the journal *Mathematics* [3]. The advantage of using these implicit RK methods lies in their symmetric or reversible character, which avoids the loss of the well-balanced property

of the methods already mentioned in the case of control techniques. This reversibility property follows from the interpretation of the methods as finding a polynomial that satisfies the ODE system at the nodes of a quadrature formula. In this paper we consider the Gauss-Legendre collocation methods whose nodes and weights coincide with those of the corresponding Gauss quadrature formulas.

This strategy has been successfully applied to a significant number of balance laws, starting with tests of lower difficulty and increasing the complexity of the problems. In particular, the following problems have been considered:

- The Burgers equation with two types of source term

$$S(U) = U^2 \quad \text{and} \quad S(U) = \sin(U).$$

- The shallow water equations (3).

- The shallow water equations with Manning friction, given by

$$\begin{cases} h_t + q_x = 0, \\ q_t + \left(\dfrac{q^2}{h} + \dfrac{1}{2}gh^2\right)_x = ghH_x - \dfrac{kq|q|}{h^\eta}, \end{cases}$$

  where k is the Manning's coefficient and η is a parameter that we take equal to $\frac{7}{3}$.

- The Euler equations with gravity.

Again, the tests show that, although designing schemes that are well-balanced is more expensive, they are more effective than standard methods when applied to simulate the propagation of small perturbations of equilibria or when long-time simulations are performed. In addition, the computational times between the different methodologies to obtain well-balanced methods have been compared. The results show that the strategy based on collocation methods is more efficient than the one developed in the works [1] and [2], in which the formulation of local problems as control problems was used.

As already mentioned, an important difficulty of these problems is due to the fact that the ODE system satisfied by the stationary solutions is not, in general, expressed in normal form. In [1] and [2], a strategy was proposed to address these problems by applying the standard reconstruction if a sonic point is detected in the stencil.

In the article [3] a general technique is introduced that allows us to work with resonant problems for systems of balance laws of the form (4). When the problem is resonant, a technique is applied in which the value of the stationary solution at a sonic point is calculated through a strategy of passing to the limit. For some systems such as shallow water, this strategy can be done analytically. In the paper, two experiments for the shallow water model are presented. In particular, the simulation of a stationary transcritical

solution and its perturbation has been considered, what allow us to verify the validity of the introduced technique.

The first three papers of this thesis are focused on the design of explicit well-balanced schemes for general systems of balance laws. The last of the papers that support this work, published by I. Gómez-Bueno, S. Boscarino, M. J. Castro, C. Parés and G. Russo in 2022 in the journal *Applied Numerical Mathematics* [4] deals with the development of high-order implicit and semi-implicit well-balanced methods. Although, in principle, it would be possible to obtain schemes of this type by applying implicit or semi-implicit numerical solvers to discretize in time the semi-discretized methods in space, the well-balanced reconstruction would lead to unfeasible problems. In particular, at each time step and in each cell, the well-balanced reconstruction process would involve the search for local equilibria related to the unknown at the next time step.

In order to avoid solving very complex linear systems, in the strategy followed in [4] we write the numerical approximation to be computed in each cell at time t^{n+1} as the sum of the one already obtained at time t^n and a fluctuation evaluated at time t^{n+1}. This fluctuation satisfies a local Cauchy problem with zero initial condition at time t^n. The method advances in time by numerically solving these Cauchy problems in the interval $[t^n, t^{n+1}]$ using implicit or semi-implicit methods. The main advantage lies in the fact that the reconstruction operators appearing in the fluctuation equations can be standard: the well-balanced property derives from the fact that, if the solution found at time t^n is a discrete stationary solution, then the solution of the Cauchy problem satisfied by the time fluctuations is identically zero. Furthermore, to reduce the computational cost and complexity of the systems, simplified reconstruction operators are used in which only the information given by the approximated solution at time t^n is used to compute the necessary coefficients and smoothness indicators.

Although we propose a general strategy to design high-order schemes, only first- and second-order methods have been numerically implemented. The proposed methods have been numerically tested considering several experiments and different systems of balance laws. Specifically, three problems are considered: the linear transport equation, the Burgers equation and the shallow water equations with topography and friction, yielding results that confirm the well-balanced character of the methods and their correct performance for the simulation of these problems with *stiff* terms.

The second part of this thesis includes the four manuscripts that support it, accompanied by a summary and a brief discussion of the obtained results. Finally, the conclusions and future work of the thesis are described. This thesis presents four works focused on the development of finite volume numerical methods for general systems of balance laws in one dimension written in the form (4), where H is a continuous function. Some possible lines of future research which are included in the section devoted to the conclusions and future work developments of the thesis are the following:

- The extension of this technique to systems of balance laws with singular source terms, i.e. with discontinuous H.

- The design of asymptotic-preserving schemes for hyperbolic systems with stiff relaxation that are also well-balanced.
- The extension of this technique to multidimensional systems.

Part I

Theoretical Framework

Chapter 1

High-order numerical methods for 1D systems of balance laws

This chapter addresses an overview of systems of balance laws, which describe a great number of relevant phenomena in fluid dynamics, such as shallow water models, multiphase flow models, gas dynamic, etc. We consider the general form of a one-dimensional hyperbolic system of balance laws:

$$U_t + f(U)_x = S(U)H_x, \tag{1.0.1}$$

where

- $U(x,t) = (u_1(x,t), ..., u_N(x,t))^T$, the vector of unknowns, takes values in an open set $\Omega \subset \mathbb{R}^N$, called set of states;
- $f : \Omega \longrightarrow \mathbb{R}^N$ is the flux function;
- $S(U)H_x$ is the source term, where $S : \Omega \longrightarrow \mathbb{R}$ and $H : \mathbb{R} \longrightarrow \mathbb{R}$ is known function (it could be the identity function).

It should be noted that (1.0.1) can be written in the quasi-linear form:

$$W_t + \mathcal{A}(W)W_x = 0, \quad x \in \mathbb{R},\, t \in [0, \infty). \tag{1.0.2}$$

Indeed, if H is taken as an unknown and the equation $H_t = 0$ is added to system (1.0.1), then one obtains (1.0.2) for the particular choices

$$W = \begin{pmatrix} U \\ H \end{pmatrix}, \quad \mathcal{A}(W) = \left(\begin{array}{c|c} J_f(U) & -S(U) \\ \hline 0 & 0 \end{array}\right),$$

where we denote by $J_f(U)$ the jacobian matrix of the flux function, that is,

$$J_f(U) = \frac{\partial f}{\partial U}(U) = \begin{pmatrix} \frac{\partial f_1}{\partial u_1}(U) & \dots & \frac{\partial f_1}{\partial u_N}(U) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial u_1}(U) & \dots & \frac{\partial f_N}{\partial u_N}(U) \end{pmatrix}. \tag{1.0.3}$$

In addition, if $S(U) = 0$ or H is a constant function, we recover the system of conservation laws of the form:

$$U_t + f(U)_x = 0, \quad x \in \mathbb{R}, t \in [0, \infty). \tag{1.0.4}$$

Again, System (1.0.4) can be rewritten in quasi-linear form as:

$$U_t + A(U)U_x = 0, \quad x \in \mathbb{R}, t \in [0, \infty), \tag{1.0.5}$$

where $A(U) = J_f(U)$.

The first section deals with the definition of weak solution for systems of the form (1.0.1) and (1.0.4): the Rankine-Hugoniot conditions, the notions of simple waves and entropy, and the solutions of Riemann problems are briefly discussed.

The second section is devoted to the design of high-order finite volume numerical methods for systems of balance laws. Some important concepts that arise when building these kind of methods are introduced. A brief overview about numerical fluxes, reconstruction operators and time integrators is included.

1.1 Weak solutions

Let us first consider sytems of the form (1.0.4). In this work, we will focus on hyperbolic systems:

Definition 1.1.1. *The system* (1.0.4) *is said to be hyperbolic if, for every $U \in \Omega$, the jacobian matrix $J_f(U)$ has N real eigenvalues*

$$\lambda_1(U) \leq ... \leq \lambda_N(U),$$

with associated eigenvectors

$$R_1(U), ..., R_N(U)$$

that generate the characteristic fields. If all the eigenvalues are distinct,

$$\lambda_1(U) < ... < \lambda_N(U),$$

the system (1.0.4) *is said to be strictly hyperbolic.*

Moreover, we suppose that the characteristic fields will be genuinely nonlinear or linearly degenerate:

Definition 1.1.2. *The characteristic field $R_i(U)$ is said to be linearly degenerate if*

$$\nabla\lambda_i(U) \cdot R_i(U) = 0, \quad \forall U \in \Omega, \tag{1.1.1}$$

where $\nabla\lambda_i(U)$ *denotes the gradient of* $\lambda_i(U)$*:*

$$\nabla\lambda_i(U) = \left(\frac{\partial\lambda_i}{\partial u_1}, \cdots, \frac{\partial\lambda_i}{\partial u_N}\right)^T.$$

The characteristic field $R_i(U)$ *is said to be genuinely nonlinear if*

$$\nabla\lambda_i(U) \cdot R_i(U) \neq 0, \quad \forall U \in \Omega. \tag{1.1.2}$$

In what follows, it is assumed that the characteristic fields are either linearly degenerate or genuinely nonlinear.

We are interested in solving Cauchy problems of the form:

$$\begin{cases} U_t + f(U)_x = 0, \\ U(x,0) = U_0(x). \end{cases} \tag{1.1.3}$$

A C^1 function $U : \mathbb{R} \times [0,\infty) \longrightarrow \Omega$ that satisfies the Cauchy problem (1.1.3) is said to be a classical solution of (1.1.3). The main issue is that, even when the initial condition U_0 is smooth, the Cauchy problems (1.1.3) may not have a classical solution. In order to address this problem, the concept of weak solution is introduced, which will be defined in terms of variational formulation.

Definition 1.1.3. *Let us consider a function* $U \in \mathcal{L}^\infty_{loc}(\mathbb{R} \times [0,\infty))$*, where* $\mathcal{L}^\infty_{loc}$ *denotes the space of locally bounded measurable functions, and the Cauchy problem* (1.1.3) *with initial condition* $U_0 \in \mathcal{L}^\infty_{loc}(\mathbb{R})$*. We say that* U *is a weak solution of* (1.1.3) *if the following conditions are fulfilled:*

- $U \in \Omega$ *almost everywhere;*
- *for any* C^1 *function* Φ *with compact support in* $\mathbb{R} \times [0,\infty)$*, one has*

$$\int_0^\infty \int_{\mathbb{R}} \Big(U(x,t)\Phi_t(x,t) + f(U(x,t))\Phi_x(x,t)\Big)\, dxdt + \int_{\mathbb{R}} U_0(x)\Phi(x,0)\, dx = 0. \tag{1.1.4}$$

If U is a weak solution of (1.0.4), then it satisfies (1.0.4) in the sense of distributions. Let us suppose that $U : \mathbb{R} \times [0,\infty) \longrightarrow \Omega$ is a piecewise C^1 function, i.e., a function which is C^1 except in a finite number of smooth curves Γ_i, $i = 1, \ldots, l$, in which U has jump discontinuities. The following result can be proved (see [85]):

Theorem 1.1.1. *A piecewise* C^1 *function* $U : \mathbb{R} \times [0,\infty) \longrightarrow \Omega$ *is a weak solution of the Cauchy problem* (1.1.3) *if and only if:*

- U *is a classical solution of* (1.1.3) *where* U *is* C^1*;*

- *the condition*

$$\sigma(U^+ - U^-) = f(U^+) - f(U^-) \tag{1.1.5}$$

is fulfilled along all the jump discontinuities, where σ is the speed of the propagation of the discontinuity and U^-, U^+ are, respectively, the left and right limit states.

The equality (1.1.5) is the so-called Rankine-Hugoniot conditions, that are usually written as

$$\sigma[U] = [f(U)], \tag{1.1.6}$$

where

$$[U] = U^+ - U^-, \quad [f(U)] = f(U^+) - f(U^-).$$

It can be shown that a piecewise C^1 function $U : \mathbb{R} \times [0, \infty) \longrightarrow \Omega$ is a weak solution if and only if

$$\int_a^b U(x, t_1)\, dx = \int_a^b U(x, t_0)\, dx + \int_{t_0}^{t_1} f(U(a, t))\, dt - \int_{t_0}^{t_1} f(U(b, t))\, dt \tag{1.1.7}$$

is satisfied for every space-time rectangle $[a, b] \times [t_0, t_1] \subset \mathbb{R} \times [0, \infty)$.

Since, in general, a weak solution for (1.1.3) is not unique, an entropy condition is required to choose, among all the possible solutions, those which have a physical meaning. In particular, it is widely extended to consider an entropy pair (η, q), where $\eta : \Omega \longrightarrow \mathbb{R}$ is a convex function, called entropy, and $q : \Omega \longrightarrow \mathbb{R}$ is a smooth function, called entropy flux, such that

$$\nabla\eta(U)^T J_f(U) = \nabla q(U)^T, \quad \forall U \in \Omega, \tag{1.1.8}$$

where

$$\nabla\eta(U) = \left(\frac{\partial \eta}{\partial u_1}, \cdots, \frac{\partial \eta}{\partial u_N}\right)^T.$$

It can be proved that classical solutions of (1.0.4) satisfy the conservation law

$$\eta(U)_t + q(U)_x = 0, \quad x \in \mathbb{R},\, t \in [0, \infty), \tag{1.1.9}$$

(see [86]). A weak solution of (1.0.4) is said to be an entropy solution if the inequality

$$\eta(U)_t + q(U)_x \leq 0 \tag{1.1.10}$$

is satisfied in the sense of distributions. Furthermore, for a piecewise C^1 weak solution U of (1.0.4), the condition (1.1.10) is fulfilled if and only if, across all the discontinuities, one has

$$\sigma[\eta(U)] \geq [q(U)]. \tag{1.1.11}$$

Alternatively, the Lax entropy condition is also widely used:

Definition 1.1.4. *Let us consider a piecewise C^1 function $U : \mathbb{R} \times [0, \infty) \longrightarrow \Omega$. A discontinuity is said to satisfy the Lax entropy conditions if there exists $i \in \{1, \cdots, N\}$ such that*

- $\lambda_i(U^-) > \sigma > \lambda_{i-1}(U^-)$ *and* $\lambda_{i+1}(U^+) > \sigma > \lambda_i(U^+)$*, if the i-th characteristic field is genuinely nonlinear;*
- $\lambda_i(U^-) = \sigma = \lambda_i(U^+)$*, if the i-th characteristic field is linearly degenerate.*

Other entropy inequalities can be also considered. Concerning the existence and uniqueness of solutions defined in the weak sense, some results can be found for initial condition close to a constant state, under a proper entropy inequality (see, for example, [87], [88], [89], [90], [91], [92]).

A class of Cauchy problems of special interest is constituted for those of the form:

$$\begin{cases} U_t + f(U)_x = 0, \\ U_0(x) = \begin{cases} U_L & \text{if } x < 0, \\ U_R & \text{if } x > 0, \end{cases} \end{cases} \tag{1.1.12}$$

where $U_L, U_R \in \Omega$, i.e., Riemann problems. The so-called elementary wave solutions or simple waves play a fundamental role in the solutions of these problems. These solutions can be classified as follows:

- Rarefaction waves, that take the form

$$U(x,t) = \begin{cases} U_L & \text{if } x/t < \lambda_i(U_L), \\ \mathcal{U}(x/t) & \text{if } \lambda_i(U_L) < x/t < \lambda_i(U_R), \\ U_R & \text{if } \lambda_i(U_R) < x/t, \end{cases} \tag{1.1.13}$$

  where $\mathcal{U}$ is a smooth function satisfying

$$J_f(\mathcal{U}(\widetilde{x}))\mathcal{U}'(\widetilde{x}) = \widetilde{x}\mathcal{U}'(\widetilde{x}), \quad \forall \widetilde{x} \in \mathbb{R}. \tag{1.1.14}$$

  These solutions are associated with genuinely nonlinear fields. Notice that the divergence condition of the characteristic

$$\lambda_i(U_L) < \lambda_i(U_R),$$

  has to be fulfilled (see [93] for details).

- Shock waves, that take the form

$$U(x,t) = \begin{cases} U_L & \text{if } x < \sigma t, \\ U_R & \text{if } x > \sigma t. \end{cases} \tag{1.1.15}$$

  These are discontinuous solutions of (1.1.12) associated with the genuinely nonlinear fields. Moreover, Rankine-Hugoniot conditions (1.1.5) and Lax entropy conditions in Definition 1.1.4 are satisfied (see [93]).

- Contact waves, that take the form

$$U(x,t) = \begin{cases} U_L & \text{if } x < \sigma t, \\ U_R & \text{if } x > \sigma t, \end{cases} \tag{1.1.16}$$

  where $\sigma = \lambda_i(U_R) = \lambda_i(U_L)$ is the propagation speed of the discontinuity for some linearly degenerate characteristic field $R_i(U)$ (see [93]).

Remark 1.1.1. *All of the types above are self-similar solutions in the sense that they can be expressed as*

$$U(x,t) = V(x/t),$$

for a function $V : \mathbb{R} \longrightarrow \Omega$.

The proof of the following result can be found in [93]:

Theorem 1.1.2. *Given a Riemann Problem* (1.1.12)*, if* U_L *and* U_R *are close enough, it has an unique weak self-similar solution that contains at most* N *simple waves of one of the three types described above.*

Let us now consider systems of the form (1.0.1). In this thesis, the function H is supposed to be continuous.

The definition of weak solution in terms of variational formulation can be naturally extended to these PDEs:

Definition 1.1.5. *Let us consider a function* $U \in \mathcal{L}^\infty_{loc}(\mathbb{R} \times [0,\infty))$*, where* $\mathcal{L}^\infty_{loc}$ *denotes the space of locally bounded measurable functions, and a Cauchy problem*

$$\begin{cases} U_t + f(U)_x = S(U)H_x, \\ U(x,0) = U_0(x). \end{cases} \tag{1.1.17}$$

with initial condition $U_0 \in \mathcal{L}^\infty_{loc}(\mathbb{R})$*. We say that* U *is a weak solution of* (1.1.3) *if the following conditions are fulfilled:*

- $U \in \Omega$ *almost everywhere;*

- *for any* C^1 *function* Φ *with compact support in* $\mathbb{R} \times [0,\infty)$*, one has*

$$\begin{aligned} &\int_0^\infty \int_{\mathbb{R}} \Big(U(x,t)\Phi_t(x,t) + f(U(x,t))\Phi_x(x,t) \Big)\, dxdt \\ &- \int_0^\infty \int_{\mathbb{R}} S(U(x,t))H_x(x)\Phi(x,t)\, dxdt + \int_{\mathbb{R}} U_0(x)\Phi(x,0)\, dx = 0. \end{aligned} \tag{1.1.18}$$

If $U : \mathbb{R} \times [0, \infty) \longrightarrow \Omega$ is a piecewise C^1 function, the Rankine-Hugoniot conditions (1.1.6) are also fulfilled along all the jump discontinuities and the necessary and sufficient condition (1.1.7) for being a weak solution is extended as:

$$\begin{aligned}\int_a^b U(x,t_1)\,dx = \int_a^b U(x,t_0)\,dx + \int_{t_0}^{t_1} f(U(a,t))\,dt - \int_{t_0}^{t_1} f(U(b,t))\,dt \\ + \int_a^b \int_{t_0}^{t_1} S(U(x,t))H_x(x)\,dtdx,\end{aligned} \tag{1.1.19}$$

for every space-time rectangle $[a,b] \times [t_0,t_1] \subset \mathbb{R} \times [0,\infty)$.

Nevertheless, the structure of the solutions of Riemann problems in the case of balance laws is more complex and will not be detailed here.

1.2 High-order finite volume numerical methods

This section is devoted to the development of high-order numerical methods for 1D systems of balance laws of the form (1.0.1). Let us consider a mesh with computational cells $I_i = [x_{i-1/2}, x_{i+1/2}]$, that is assumed to be uniform for simplicity with constant length $\Delta x = x_{i+1/2} - x_{i-1/2}$. Here, $x_{i+1/2} = i\Delta x$ is the intercell and the centre of the cell is represented by $x_i = \left(i - \frac{1}{2}\right)\Delta x$.

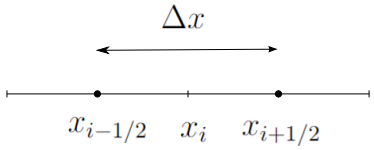


Figure 1.1: Sketch of a computational cell I_i.

1.2.1 High-order finite volume numerical methods for systems of conservation laws

Let us first consider systems of conservation laws of the form (1.0.4). Let $U(x,t)$ denote a solution of (1.0.4). Averaging (1.0.4) over every cell I_i, one has

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U_t\,dx = -\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} f(U)_x\,dx. \tag{1.2.1}$$

If we now use $\widetilde{U}_i(t)$ to symbolise the approximation of the average of the solution $U(x,t)$ in the cell I_i at time t

$$\widetilde{U}_i \approx \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U\,dx, \tag{1.2.2}$$

we obtain semi-discrete conservative numerical methods of the form:

$$\frac{d\widetilde{U}_i(t)}{dt} = -\frac{1}{\Delta x}\left(F^t_{i+1/2} - F^t_{i-1/2}\right), \tag{1.2.3}$$

where $F^t_{i+1/2}$ is an approximation of the average of the flux at $x_{i+1/2}$.

Definition 1.2.1. *A numerical method is said to be conservative if it can be written in the form* (1.2.6) *with*

$$F^t_{i+1/2} = \mathbb{F}(\widetilde{U}_{i-q}(t), \ldots, \widetilde{U}_{i+p}(t)),$$

where $\mathbb{F}$ *is a Lipschitz continuous function, called numerical flux, such that*

$$\mathbb{F}(U, \ldots, U) = f(U), \quad \forall U \in \Omega.$$

In order to obtain high-order numerical methods, we will consider high-order reconstruction operators.

Definition 1.2.2. *A consistent operator* P_i *is said to be a reconstruction operator of order* p *if, given a sequence of cell values* $\{\widetilde{U}_i\}$*, it provides a smooth function at every cell* I_i *by means of the values in some neighbour cells belonging to the stencil of the* i*-th cell:*

$$P_i(x) = P_i(x; \{\widetilde{U}_j\}_{j\in\mathcal{S}_i}),$$

where $\mathcal{S}_i$ *denotes the set of indexes of the stencil. Additionally, the operator* P_i *must fulfil the following two properties:*

- *Conservation property:*

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} P_i(x)\,dx = \widetilde{U}_i, \quad \forall i. \tag{1.2.4}$$

- *Approximation property:*
  Let us suppose that $\{\widetilde{U}_i\}$ *is the vector of cell averages of a smooth function* $U(x)$*:*

$$\widetilde{U}_i = \int_{x_{i-1/2}}^{x_{i+1/2}} U(x)\,dx.$$

  Then, there exists $p \in \mathbb{N}$ *such that*

$$P_i(x) = U(x) + O(\Delta x^p), \quad \forall x \in [x_{i-1/2}, x_{i+1/2}]. \tag{1.2.5}$$

Therefore, we consider semi-discrete high-order numerical methods of the form:

$$\frac{d\widetilde{U}_i(t)}{dt} = -\frac{1}{\Delta x}\left(F^t_{i+1/2} - F^t_{i-1/2}\right), \tag{1.2.6}$$

where

$$F^t_{i+1/2} = \mathbb{F}(U^{t,-}_{i+1/2}, U^{t,+}_{i+1/2}).$$

Here $\mathbb{F}$ is a consistent numerical flux, which is evaluated at the so-called reconstructed states at the intercells $U^{t,\pm}_{i+1/2}$ that we compute by means of a high-order reconstruction operator as follows:

$$U^{t,-}_{i+1/2} = P^t_i(x_{i+1/2}), \quad U^{t,+}_{i+1/2} = P^t_{i+1}(x_{i+1/2}),$$

where $P^t_i(x)$ is a reconstruction of the solution at the cell I_i at time t with order of accuracy p, that is defined in terms of the vector of cell averages $\{\widetilde{U}_i(t)\}$:

$$P^t_i(x) = P_i(x; \{\widetilde{U}_j(t)\}_{j\in\mathcal{S}_i}). \tag{1.2.7}$$

$\mathcal{S}_i$ stands for the set of indices belonging to the stencil linked to the reconstruction operator $P^t_i(x)$.

Please note that, since H is a continuous function and differentiable almost everywhere, only two ingredients are enough to develop high-order numerical methods: a consistent numerical flux and a high-order reconstruction operator.

1.2.2 Numerical fluxes

Let us highlight some well-known consistent numerical fluxes $\mathbb{F}$ that have been applied in this work (see, for example, [86], [93], [15]).

- Lax-Friedrichs numerical flux:

$$\mathbb{F}(U_L, U_R) = \frac{f(U_L) + f(U_R)}{2} - \frac{\Delta x}{2\Delta t}(U_R - U_L). \tag{1.2.8}$$

  The corresponding first-order numerical scheme associated to this flux is stable under the usual CFL condition

$$\frac{\Delta t}{\Delta x} \max_{l=1,\dots,M} \left(\max_{j=1,\dots,N} |\lambda_j(U_l)| \right) = \text{CFL},$$

  where M is the number of computational of cells and CFL$\in (0, 1]$.

- Local Lax-Friedrichs or Rusanov numerical flux, which is a variation of the previous one where $\dfrac{\Delta x}{\Delta t}$ is replaced by a local value:

$$\mathbb{F}(U_L, U_R) = \frac{f(U_L) + f(U_R)}{2} - \frac{\alpha(U_L, U_R)}{2}(U_R - U_L), \tag{1.2.9}$$

  where

$$\alpha(U_L, U_R) = \max\{|\lambda_j(U_L)|, |\lambda_j(U_R)|; j = 1, \dots, N\}.$$

- HLL numerical flux:

$$\mathbb{F}(U_L,U_R)=\begin{cases} f(U_L) & \text{if } 0\leq S_L,\\ \dfrac{S_Rf(U_L)-S_Lf(U_R)+S_LS_R(U_R-U_L)}{S_R-S_L} & \text{if } S_L<0<S_R,\\ f(U_R) & \text{if } 0\geq S_R, \end{cases} \tag{1.2.10}$$

where

$$S_L=\min\{\lambda_1(U_L),\lambda_1(U_R)\},\quad S_R=\max\{\lambda_N(U_L),\lambda_N(U_R)\}.$$

An alternative expression in compact form is given by:

$$\begin{aligned}\mathbb{F}(U_L,U_R)=&\left(f(U_L)(S_R+|S_R|-S_L-|S_L|)+f(U_R)(S_R-|S_R|-S_L+|S_L|)\right.\\ &\left.-(S_R|S_L|-S_L|S_R|)(U_R-U_L)\right)/2(S_R-S_L).\end{aligned} \tag{1.2.11}$$

Again, the corresponding method is stable under the usual CFL condition.

1.2.3 Reconstruction operators

In the design of high-order numerical methods of order p, reconstruction operators play a crucial role. This key ingredient constitutes the basis of this work, and it is usually obtained by means of approximation or interpolation approaches.

The MUSCL, ENO, WENO and CWENO operators or the hyperbolic reconstructions are some examples of the most popular reconstruction operators (see, for instance, [94], [95], [96], [97], [98], [14], [99], [100], [101]). In this thesis, we will focus on the MUSCL and CWENO reconstructions and they are briefly introduced below. Although WENO reconstruction is not used, it is also recalled to help to clarify the choice of the CWENO one as an alternative to circumvent some downsides of the WENO reconstruction in the framework of the numerical approximation of balance laws.

1.2.3.1 MUSCL reconstruction operator

Since it is second-order accurate, this reconstruction operator has been applied in the works underlying this thesis for the design of second-order schemes:

$$U^{\pm}_{i+1/2}=U(x_{i+1/2})+O(\Delta x^2). \tag{1.2.12}$$

Given a vector of cell values $\{U_i\}$, the MUSCL reconstruction is based on a piecewise linear reconstruction of the form

$$P_i(x)=U_i+\Delta_iU(x-x_i), \tag{1.2.13}$$

where Δ_iU is an approximation of the first-order spatial derivative at x_i. Introduced in [102], its initials stand for "Monotone Upwind Scheme for Conservation Laws" (see [103],

[104] for further information). In order to obtain an approximation $\Delta_i U$, some strategies to limit the slope of (1.2.13) are required. In this work, two slope limiters have been considered to obtain $\Delta_i U$: the *minmod* and the *avg* or *harmod* limiters [105].

- The *minmod* limiter is given by

$$\Delta_i U = minmod\left(\frac{U_{i+1}-U_i}{\Delta x}, \frac{U_i - U_{i-1}}{\Delta x}\right), \tag{1.2.14}$$

where

$$minmod(a,b) = \begin{cases} \min(a,b) & if\ a>0, b>0, \\ \max(a,b) & if\ a<0, b<0, \\ 0 & otherwise. \end{cases} \tag{1.2.15}$$

- Similarly, the *avg* or *harmod* limiter is given by

$$\Delta_i U = avg\left(\frac{U_{i+1}-U_i}{\Delta x}, \frac{U_i - U_{i-1}}{\Delta x}\right), \tag{1.2.16}$$

where

$$avg(a,b) = \begin{cases} \dfrac{|a|b+|b|a}{|a|+|b|} & if\ |a|+|b|>0, \\ 0 & otherwise. \end{cases} \tag{1.2.17}$$

Note that (1.2.15) and (1.2.17) are computed componetwise.

Observe that in this case the stencil is composed of two neighbour cells. Consequently, the reconstructed states at the intercells are given by

$$\begin{aligned} U^+_{i-1/2} &= U_i + \Delta_i U(x_{i-1/2} - x_i) = U_i - \frac{\Delta x}{2}\Delta_i U, \\ U^-_{i+1/2} &= U_i + \Delta_i U(x_{i+1/2} - x_i) = U_i + \frac{\Delta x}{2}\Delta_i U. \end{aligned} \tag{1.2.18}$$

Moreover, it is obvious that P_i defined as (1.2.13) is a conservative reconstruction operator.

1.2.3.2 WENO reconstruction operator

The WENO reconstruction operator (Weighted Essentially Non Oscillatory) is based on a piecewise polynomial reconstruction from a sequence of averages $\{U_i\}$ in the cells of a function $U(x)$ (see [13], [14], [99]).

For every cell I_i, let us consider a stencil composed by $2g+1$ cells, whose set of indexes is given by:

$$\mathcal{S}_i = \{i-g, \ldots, i+g\},$$

and let us denote by P_i^{opt} the interpolation polynomial of degree $G = 2g$ that interpolates the $2g+1$ averages in the cells of the stencil $\{U_j\}_{j\in\mathcal{S}_i}$. Due to the fact that P_i^{opt} may present

spurious oscillations if a discontinuity exists in the stencil, in the WENO reconstruction instead of directly using P_i^{opt}, polynomials of lower degree g are computed to avoid the discontinuities.

Firstly, let us show how to build P_i^{opt}. Let us denote by $\mathcal{U}(x)$ a primitive of the function $U(x)$:

$$\mathcal{U}(x) = \int_{-\infty}^{x} U(\xi)\, d\xi.$$

Observe that $\mathcal{U}(x_{i+1/2})$ can be defined in terms of the cell averages of $U(x)$ as follows:

$$\mathcal{U}(x_{i+1/2}) = \sum_{j=-\infty}^{i} \int_{x_{j-1/2}}^{x_j+1/2} U(\xi)\, d\xi = \sum_{j=-\infty}^{i} \Delta x U_j.$$

Thus, given the cell averages $\{U_i\}$, the values of the primitive $\mathcal{U}(x)$ at the intercells is known. If $\widehat{P_i^{opt}}$ is the interpolation polynomial of degree G that interpolates the primitive $\mathcal{U}$ evaluated at the intercell of the cells belonging to the stencil $\mathcal{S}_i$, and P_i^{opt} denotes its derivative

$$P_i^{opt}(x) = \widehat{P_i^{opt}}',$$

one has

$$\begin{aligned}
\frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} P_i^{opt}(\xi)\, d\xi &= \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \widehat{P_i^{opt}}'(\xi)\, d\xi \\
&= \frac{1}{\Delta x}\left(\widehat{P_i^{opt}}(x_{j+1/2}) - \widehat{P_i^{opt}}(x_{j-1/2})\right) \\
&= \frac{1}{\Delta x}\left(\mathcal{U}(x_{j+1/2}) - \mathcal{U}(x_{j-1/2})\right) \\
&= \frac{1}{\Delta x}\left(\int_{-\infty}^{x_{j+1/2}} U(\xi)\, d\xi - \int_{-\infty}^{x_{j-1/2}} U(\xi)\, d\xi\right) \\
&= \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} U(\xi)\, d\xi = U_j, \quad j \in S_i,
\end{aligned} \tag{1.2.19}$$

where the fact that $\widehat{P_i^{opt}}$ interpolates $\mathcal{U}$ at the intercells has been applied in the third equality.

To continue with the description of the WENO reconstruction, let us denote by $P_i^k, k = 1, \ldots, g+1$ the interpolation polynomial of degree g that interpolates the averages of the cells belonging to the substecils $\mathcal{S}_i^k$ whose set of indexes are given by

$$\mathcal{S}_i^k = \{i - g + k - 1, \ldots, i + k - 1\}, \quad k = 1, \ldots, g+1.$$

The expression of the WENO reconstruction operator depends on the polynomials P_i^k, $k = 1, \ldots, g+1$ and P_i^{opt}:

$$P_i(\hat{x}) = \mathrm{WENO}(P_i^1, \ldots, P_i^{g+1}; P_i^{opt}, \hat{x}), \quad \hat{x} \in [x_{i-1/2}, x_{i+1/2}]. \tag{1.2.20}$$

This operator also depends on the choice of a set of coefficients, known as linear coefficients, as follows:

- The first step to build the WENO reconstruction operator is to look for the set of linear coefficients $d_i^1(\hat{x}), \ldots, d_i^{g+1}(\hat{x})$, also called optimal coefficients, that verify:

$$\sum_{k=1}^{g+1} d^k(\hat{x}) P_i^k(\hat{x}) = P_i^{opt}(\hat{x}), \quad \sum_{k=1}^{g+1} d_i^k(\hat{x}) = 1. \tag{1.2.21}$$

- Secondly, the nonlinear coefficients $w_i^1(\hat{x}), \ldots, w_i^{g+1}(\hat{x})$ are computed from the linear ones by means of the formulas:

$$\alpha_i^k(\hat{x}) = \frac{d_i^k(\hat{x})}{(I[P_i^k] + \epsilon)^\mu}, \quad w_i^k(\hat{x}) = \frac{\alpha_i^k(\hat{x})}{\sum_{j=1}^{g+1} \alpha_i^j(\hat{x})}, \quad k = 1, \ldots, g+1, \tag{1.2.22}$$

  where

  - $I[P_i^k]$ is an smoothness indicator evaluated at P_i^k. We consider the well-known Jiang-Shu indicator

$$I[P_i^k] = \sum_{l \geq 1} diam(I_i)^{2l-1} \int_{I_i} \left(\frac{d^l}{dx^l} P_i^k(x) \right) dx, \tag{1.2.23}$$

    being $\frac{d^l}{dx^l} P_i^k(x)$ the l-th derivative of P_i^k (see [106]). Some other examples of smoothness indicators can be found in [107], [108];
  - ϵ is a small positive value;
  - $\mu \geq 2$.

Finally, the WENO reconstruction operator at $\hat{x} \in I_i$ is defined as follows:

$$P_i(\hat{x}) = \sum_{k=1}^{p+1} w_i^k(\hat{x}) P_i^k(\hat{x}). \tag{1.2.24}$$

Be aware that an important drawback of the WENO reconstruction operator is the fact that the coefficients depend on each point $\hat{x} \in I_i$ explicitly. What is more, for interior points, the linear coefficients may not exist or may be non-positive. The CWENO reconstruction, which will be introduced below, allows us to circumvent this difficulty.

1.2.3.3 CWENO reconstruction operator

We introduce now the CWENO reconstruction operator, which allows us to remedy some of the highlighted disadvantages of the WENO reconstruction. In this case, the linear coefficients are rather arbitrarily chosen and, consequently, they can be uniform for every point in the cell. As a result, it is enough to compute the values α_i^k and the non-linear coefficients w_i^k only once at each cell.

The CWENO reconstruction operator is given by

$$P_i = \text{CWENO}(P_i^1, \ldots, P_i^m; P_i^{opt}), \tag{1.2.25}$$

where $P_i^1, \ldots, P_i^m$ are a set of interpolation polynomials of degree $g < G$ with $m \geq 1$ and P_i^{opt} is again the optimal polynomial of degree G that interpolates the $G+1$ cell averages belonging to the stencil. Now, the CWENO reconstruction operator depends on the choice of a set of linear coefficients $d_i^0, d_i^1, \ldots, d_i^m$ satisfying

$$d_i^k \in [0,1],\ \forall k \in \{0,1,\ldots,m\}, \quad \sum_{k=0}^{m} d_i^k = 1, \quad d_i^0 \neq 0, \tag{1.2.26}$$

as follows:

- First, compute the polynomial P_i^0 of degree G with expression

$$P_i^0(x) = \frac{1}{d_i^0}\left(P_i^{opt}(x) - \sum_{k=1}^{m} d_i^k P_i^k(x)\right). \tag{1.2.27}$$

- Next, compute the nonlinear coefficients w_i^k from the linear ones:

$$\alpha_i^k = \frac{d_i^k}{(I[P_i^k] + \epsilon)^\mu}, \quad w_i^k = \frac{\alpha_i^k}{\sum_{j=0}^{m} \alpha_i^j}, \quad k = 0, \ldots, m. \tag{1.2.28}$$

Finally, the CWENO reconstruction operator is defined as:

$$P_i(x) = \sum_{k=0}^{m} w_i^k P_i^k(x). \tag{1.2.29}$$

In this thesis, we will focus on CWENO3, i.e., the CWENO reconstruction operator of order three. If we particularise the description above to the third order case, the CWENO3 reconstruction operator is given by:

$$P_i(x) = \sum_{k=0}^{2} w_i^k P_i^k(x), \tag{1.2.30}$$

where

- P_i^{opt} is the polynomial of degree 2 that interpolates $\{U_{i-1}, U_i, U_{i+1}\}$;
- P_i^1 is the polynomial of degree 1 that interpolates $\{U_{i-1}, U_i\}$;
- P_i^2 is the polynomial of degree 1 that interpolates $\{U_i, U_{i+1}\}$;
- P_i^0 the polynomial that fulfils

$$P_i^0(x) = \frac{1}{d_i^0}\left(P_i^{opt}(x) - d_i^1 P_i^1(x) - d_i^2 P_i^2(x)\right);$$

- the linear coefficients meet

$$d_i^k \in [0,1],\ \forall k \in \{0,1,2\}, \quad \sum_{k=0}^{2} d_i^k = 1, \quad d_i^0 \neq 0; \tag{1.2.31}$$

- the non-linear coeffients are obtained as

$$w_i^k = \frac{\alpha_i^k}{\sum_{j=0}^{2}\alpha_i^j}, \quad \text{with} \quad \alpha_i^k = \frac{d_i^k}{(I[P_i^k]+\epsilon)^\mu}, \quad k = 0,1,2. \tag{1.2.32}$$

We have taken the particular choices

$$d_i^0 = 0.7, \quad d_i^1 = 0.15, \quad d_i^2 = 0.15$$

for the linear coefficients. The CWENO3 reconstruction has been also explored in [109] and [110] with another choice of the linear coefficients.

1.2.4 High-order finite volume numerical methods for systems of balance laws

Let us now consider a systems of balance laws of the form (1.0.1). The standard form of a semi-discrete high-order numerical methods writes as follows:

$$\frac{d\widetilde{U}_i(t)}{dt} = -\frac{1}{\Delta x}\left(F_{i+1/2}^t - F_{i-1/2}^t\right) + \frac{1}{\Delta x} S_i^t, \tag{1.2.33}$$

where the numerical fluxes are defined as in (1.2.6) and S_i^t is the approximation of the integral of the source term at the i-th cell, i.e.,

$$S_i^t \approx \int_{x_{i-1/2}}^{x_{i+1/2}} S(P_i^t(x)) H_x \, dx. \tag{1.2.34}$$

At this step all the ingredients to obtain semi-discrete in space high-order numerical methods have been defined and the only remaining requirement is to evaluate (1.2.33)-(1.2.34) with the chosen reconstruction operator.

1.2.5 Time discretization

Notice that the semi-discrete numerical method (1.2.33) is an ODE system that must be solved by applying a numerical solver with the desired order. We denote by Δt the time step and by t^n the n-th time.

1.2.5.1 Explicit time discretization

If an explicit ODE solver is applied to (1.2.33), fully-discrete methods are obtained. In this thesis, we consider the explicit total variation diminishing (TVD) Runge-Kutta methods from [12] and [14]: let us rewrite (1.2.33) in the form

$$\frac{d\widetilde{U}_i(t)}{dt} = L(\widetilde{U}_i(t)), \tag{1.2.35}$$

where $L(U)$ represents the right-hand side. In particular, we focus on first-, second- and third-order in time schemes whose expression are the following:

- First-order:

$$\widetilde{U}_i^{n+1} = \widetilde{U}_i^n - \Delta t L(\widetilde{U}_i^n). \tag{1.2.36}$$

- Second-order:

$$\begin{cases} \widetilde{U}_i^{n+1/2} = \widetilde{U}_i^n - \Delta t L(\widetilde{U}_i^n), \\ \widetilde{U}_i^{n+1} = \dfrac{1}{2}\widetilde{U}_i^n + \dfrac{1}{2}\left(\widetilde{U}_i^{n+1/2} - \Delta t L\left(\widetilde{U}_i^{n+1/2}\right)\right). \end{cases} \tag{1.2.37}$$

- Third-order:

$$\begin{cases} \widetilde{U}_i^{n+1/3} = \widetilde{U}_i^n - \Delta t L(\widetilde{U}_i^n), \\ \widetilde{U}_i^{n+2/3} = \dfrac{3}{4}\widetilde{U}_i^n + \dfrac{1}{4}\left(\widetilde{U}_i^{n+1/3} - \Delta t L\left(\widetilde{U}_i^{n+1/3}\right)\right), \\ \widetilde{U}_i^{n+1} = \dfrac{1}{3}\widetilde{U}_i^n + \dfrac{2}{3}\left(\widetilde{U}_i^{n+2/3} - \Delta t L\left(\widetilde{U}_i^{n+2/3}\right)\right). \end{cases} \tag{1.2.38}$$

The ADER method is an alternative to reach high-order both in time and space in a single step. Introduced by Toro and Titarev in the papers [111], [112], [113], [114], this technique has been widely explored in the literature: see, for instance, [115], [116], [96], [117], [118], [119], [120], [121], [122]. In [123] another alternative was introduced based on the approximation of the derivatives in the Taylor expansion in time of the solution through high-order central divided difference formulas in a recursive way. A modification of this strategy leading to numerical methods using stencils of minimal length was introduced in [124]. In both cases, WENO reconstructions are used to avoid spurious oscillations: see also [108].

1.2.5.2 Implicit time discretization

Let us consider the particular case where both the hyperbolic term and the source of (1.0.1) are stiff. An implicit treatment of (1.2.35) is then required. In this work, the so-called diagonally implicit Runge-Kutta (DIRK) schemes and, in particular, the *singly diagonally implicit* with Butcher tableau

$$\begin{array}{c|ccccc} \gamma & \gamma & 0 & 0 & \dots & 0 \\ c_2 & a_{2,1} & \gamma & 0 & \dots & 0 \\ c_3 & a_{3,1} & a_{3,2} & \gamma & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{s,1} & a_{s,2} & a_{s,3} & \dots & \gamma \\ \hline & b_1 & b_2 & b_3 & \dots & b_s \end{array} \tag{1.2.39}$$

have been considered. The application of a DIRK method (1.2.39) to (1.2.35) leads to a numerical method of the form

$$\begin{aligned} \widetilde{U}_i^k &= \widetilde{U}_i^n + \Delta t \sum_{l=1}^{k-1} a_{k,l} L(\widetilde{U}_i^l) + \Delta t \gamma L(\widetilde{U}_i^k), \quad k = 1, \dots, s, \\ \widetilde{U}_i^{n+1} &= \widetilde{U}_i^n + \Delta t \sum_{l=1}^{s} b_l L(\widetilde{U}_i^l). \end{aligned}$$

If the scheme is is *stiffly accurate*, that is,

$$a_{s,i} = b_i, \quad i = 1, \dots, s,$$

the numerical solution at time t^{n+1} agrees with the last stage value

$$\widetilde{U}_i^{n+1} = \widetilde{U}_i^s.$$

By way of illustration, let us consider the second-order DIRK method with Butcher tableau

$$\begin{array}{c|cc} \gamma & \gamma & 0 \\ 1 & 1-\gamma & \gamma \\ \hline & 1-\gamma & \gamma, \end{array} \tag{1.2.40}$$

where $\gamma = 1 - \dfrac{1}{\sqrt{2}}$ for the time discretization. The fully discrete implicit numerical method writes:

$$\begin{aligned} \widetilde{U}_i^1 &= \widetilde{U}_i^n + \Delta t \gamma L(\widetilde{U}_i^1), \\ \widetilde{U}_i^{n+1} &= \widetilde{U}_i^n + \frac{(1-\gamma)}{\gamma} \widetilde{U}_i^1 + \Delta t \gamma L(\widetilde{U}_i^{n+1}). \end{aligned} \tag{1.2.41}$$

1.2.5.3 Semi-implicit time discretization

Let us assume now that (1.0.1) can be written as a problem of the form

$$U_t + f^1(U)_x + f^2(U)_x = S^1(U)H_x + S^2(U), \tag{1.2.42}$$

with f^1 and S^1 non stiff, and f^2 and S^2 stiff. Here, the functions f^i, S^i, $i = 1, 2$ can vanish. Then, in this case, the numerical method in semi-discrete version takes the form

$$\frac{d\widetilde{U}_i(t)}{dt} = L^1(\widetilde{U}_i(t)) + L^2(\widetilde{U}_i(t)), \tag{1.2.43}$$

where $L^1(U)$ is a spatial discretization operator for the non stiff terms, and $L^2(U)$ for the stiff ones.

An IMEX method can be adopted treating the non stiff terms explicitly while an implicit treatment of the stiff terms is performed, with a double Butcher tableau of the form

$$\begin{array}{c|ccccc} 0 & 0 & 0 & 0 & \dots & 0 \\ \tilde{c}_2 & \tilde{a}_{2,1} & 0 & 0 & \dots & 0 \\ \tilde{c}_3 & \tilde{a}_{3,1} & \tilde{a}_{3,2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{c}_s & \tilde{a}_{s,1} & \tilde{a}_{s,2} & \tilde{a}_{s,3} & \dots & 0 \\ \hline & \tilde{b}_1 & \tilde{b}_2 & \tilde{b}_3 & \dots & \tilde{b}_s \end{array}, \qquad \begin{array}{c|ccccc} \gamma & \gamma & 0 & 0 & \dots & 0 \\ c_2 & a_{2,1} & \gamma & 0 & \dots & 0 \\ c_3 & a_{3,1} & a_{3,2} & \gamma & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{s,1} & a_{s,2} & a_{s,3} & \dots & \gamma \\ \hline & a_{s,1} & a_{s,2} & a_{s,3} & \dots & \gamma \end{array} \tag{1.2.44}$$

(see, for instance, [17], [19]). Then, the fully discrete numerical methods reads as follows:

$$\begin{aligned} \widetilde{U}_i^k &= \widetilde{U}_i^n + \Delta t \sum_{l=1}^{k-1} \tilde{a}_{k,l} L^1(\widetilde{U}_i^l) + \Delta t \sum_{l=1}^{k-1} a_{k,l} L^2(\widetilde{U}_i^l) + \Delta t \gamma L^2(\widetilde{U}_i^k), \quad k = 1, \dots, s, \\ \widetilde{U}_i^{n+1} &= \widetilde{U}_i^n + \Delta t \sum_{l=1}^{s} \tilde{b}_i L^1(\widetilde{U}_i^l) + \Delta t \sum_{l=1}^{s-1} a_{s,i} L^2(\widetilde{U}_i^l) + \Delta t \gamma L^2(\widetilde{U}_i^s). \end{aligned}$$

As it has been done for the implicit case, let us illustrate this scheme with the second-order IMEX method with tableaux

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2\gamma} & \frac{1}{2\gamma} & 0 \\ \hline & 1-\gamma & \gamma \end{array}, \qquad \begin{array}{c|cc} \gamma & \gamma & 0 \\ 1 & 1-\gamma & \gamma \\ \hline & 1-\gamma & \gamma \end{array}, \tag{1.2.45}$$

with $\gamma = 1 - \dfrac{1}{\sqrt{2}}$. The stages of the fully discrete numerical method are:

$$\begin{aligned} \widetilde{U}_i^1 &= \widetilde{U}_i^n + \Delta t \gamma L^2(\widetilde{U}_i^1), \\ \widetilde{U}_i^2 &= \widetilde{U}_i^n + \frac{\Delta t}{2\gamma} L^1(\widetilde{U}_i^1) + \frac{1-\gamma}{\gamma} \widetilde{U}_i^1 + \Delta t \gamma L^2(\widetilde{U}_i^2), \\ \widetilde{U}_i^{n+1} &= \widetilde{U}_i^n + \widetilde{U}_i^2 + \Delta t \left(1 - \gamma - \frac{1}{2\gamma}\right) L^1(\widetilde{U}_i^1) + \gamma \Delta t L^1(\widetilde{U}_i^2). \end{aligned}$$

Chapter 2

Exactly well-balanced numerical methods

We consider systems of balance laws of the form (1.0.1), which admit non-trivial stationary solutions satisfying

$$f(U)_x = S(U)H_x. \tag{2.0.1}$$

These systems are widely used for the simulation of the waves generated by small perturbations of steady states: think, for instance, of tsunami waves in the ocean. A numerical method with the property of solving exactly or with enhanced accuracy all the steady states of (1.0.1) or, at least, a relevant family of them is said to be *exactly well-balanced* (EWB). What is more, in the first instance, the method is defined as *fully exactly well-balanced*.

If standard methods are applied to (1.0.1) with an initial condition that represents a perturbation of a stationary solution, the discretization errors perturb the underlying steady state all over the computational domain from the first iteration in time. If these errors are of the same order of magnitude of the initial perturbation, distinguishing between the waves to be simulated and those which appear as a result of the discretization errors is not possible. Moreover, although the discretization errors can be reduced by refining the mesh or by considering higher order methods, the growth of the computational cost may not be affordable. Bermudez and Vázquez-Cendón introduced in [78] the *C-property* condition in the context of the shallow water equations referring to schemes which exactly preserve the stationary solutions corresponding to water at rest. In this thesis, exactly well-balanced schemes will be developed by following the strategy described by Castro and Parés in [32], which is based on the use of reconstruction operators that are well-balanced in the following sense:

Definition 2.0.1. *Given a stationary solution U^* of* (1.0.1) *a reconstruction operator is said to be exactly well-balanced for U^* if, for every cell index i, the following equality holds:*

$$P_i(x) = U^*(x), \quad \forall x \in I_i,$$

where P_i is the approximation of U^* obtained by applying the reconstruction operator to the sequence of cell averages $\{\overline{U}_i^*\}$ of U^*:

$$\overline{U}_i^* = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U^*(x)\,dx.$$

Once the concept of exactly well-balanced reconstruction operator has been introduced, the following result can be proved:

Theorem 2.0.1. *If the reconstruction operator P_i is exactly well-balanced for a continuous stationary solution U^* of* (1.0.1)*, then the numerical method* (1.2.33) *with*

$$S_i^t = \int_{x_{i-1/2}}^{x_{i+1/2}} S(P_i^t(x))H_x(x)\,dx, \tag{2.0.2}$$

is also exactly well-balanced for U^*, i.e., the sequence of cell averages of U^* is an equilibrium of the ODE given by the numerical method (1.2.33)-(2.0.2)*, where P_i^t is defined as* (1.2.7)*.*

Proof. Let U^* be a stationary solution of (1.0.1) and let P_i^0 be an exactly well-balanced reconstruction operator applied to the vector $\{\overline{U}_i^*\}$ of exact cell averages of U^*. Let us show that $\{\overline{U}_i^*\}$ is an equilibrium of (1.2.33)-(2.0.2).

Considering that P_i^0 is an exactly well-balanced reconstruction operator, one has

$$\begin{aligned} P_i^0 &= U^*(x), \quad \forall x \in I_i,\ \forall i,\\ U_{i+1/2}^{0,-} &= P_i^0(x_{i+1/2}) = U^*(x_{i+1/2}),\\ U_{i+1/2}^{0,+} &= P_{i+1}^0(x_{i+1/2}) = U^*(x_{i+1/2}). \end{aligned}$$

Therefore

$$\begin{aligned} &F_{i+1/2}^0 - F_{i-1/2}^0 - \int_{x_{i-1/2}}^{x_{i+1/2}} S(P_i^0(x))H_x(x)\,dx\\ &\quad= \mathbb{F}(U_{i+1/2}^{0,-}, U_{i+1/2}^{0,+}) - \mathbb{F}(U_{i-1/2}^{0,-}, U_{i-1/2}^{0,+}) - \int_{x_{i-1/2}}^{x_{i+1/2}} S(U^*(x))H_x(x)\,dx\\ &\quad= \mathbb{F}(U^*(x_{i+1/2}), U^*(x_{i+1/2})) - \mathbb{F}(U^*(x_{i-1/2}), U^*(x_{i-1/2})) - \int_{x_{i-1/2}}^{x_{i+1/2}} S(U^*(x))H_x(x)\,dx\\ &\quad= f(U^*(x_{i+1/2})) - f(U^*(x_{i-1/2})) - \int_{x_{i-1/2}}^{x_{i+1/2}} S(U^*(x))H_x(x)\,dx = 0, \end{aligned}$$

where in the second-to-last equality the consistency of the numerical flux is used, whereas in the last equality we have applied the fact that U^* is a stationary solution, hence, it is a solution of the ODE (2.0.1). Then, the family $\{\overline{U}_i^*\}$ of cell averages of U^* is an equilibrium of (1.2.33)-(2.0.2), as we wanted to show. □

2.1 Exactly well-balanced reconstruction procedure

The main difficulty when designing exactly well-balanced methods is the fact that, in general, standard reconstruction operators such as the ones introduced in Chapter 1 (MUSCL, WENO, CWENO,...), that will be denoted in this section as

$$Q_i(x) = Q_i(x; \{\overline{U}_j\}_{j\in\mathcal{S}_i}),$$

are not expected to be exactly well-balanced. We will follow the methodology presented in [125] and developed in [32] to obtain exactly well-balanced reconstruction operators from the basis of standard ones:

Algorithm 2.1.1. *Given a family of cell values $\{\overline{U}_i\}$, at every cell I_i:*

1. *Find, if possible, a stationary solution $U_i^*(x)$ in the cells belonging to the stencil of I_i, $\cup_{j\in\mathcal{S}_i} I_j$, such that:*
$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U_i^*(x)\,dx = \overline{U}_i. \tag{2.1.1}$$
Otherwise, take $U_i^* \equiv 0$.

2. *Apply a standard reconstruction operator of order p to the cell values $\{V_j\}_{j\in\mathcal{S}_i}$, called fluctuations, given by*
$$V_j = \overline{U}_j - \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} U_i^*(x)\,dx, \quad j\in\mathcal{S}_i,$$
to obtain:
$$Q_i(x) = Q_i(x; \{V_j\}_{j\in\mathcal{S}_i}).$$

3. *Finally, define*
$$P_i(x) = U_i^*(x) + Q_i(x). \tag{2.1.2}$$

Remark 2.1.1. *Notice that if the first step has more than one solution, a criterion to choose one of them is required. For example, in [79] this issue was tackled for the shallow water equations.*

The following result is satisfied by the reconstruction operator defined in (2.1.2):

Theorem 2.1.1. *The reconstruction operator P_i in* (2.1.2) *is exactly well-balanced for any stationary solution of* (1.0.1) *provided that Q_i is exact for the zero function. Additionally, P_i is conservative provided that Q_i is conservative and it is high-order accurate provided that the stationary solutions are smooth.*

Proof. Let us first check that P_i is exactly well-balanced for any stationary solution. Let U^* be a stationary solution of (1.0.1) and let us consider the vector $\{\overline{U}_i^*\}$ of exact cell averages of U^*:

$$\overline{U}_i^* = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U^*(x)\,dx.$$

Then, the stationary solution U_i^* found in the first step of the reconstruction procedure at every cell I_i is

$$U_i^*(x) = U^*(x), \quad \forall x \in I_i,$$

what implies

$$V_j = \overline{U}_j^* - \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} U^*(x)\,dx = 0, \quad j \in \mathcal{S}_i,\ \forall t,$$

i.e., the fluctuations $\{V_j\}_{j\in\mathcal{S}_i}$ are equal to zero. If Q_i is exact for the zero function, one has

$$P_i(x) = U^*(x), \quad \forall x \in I_i,\ \forall i,$$

that is, P_i is an exactly well-balanced reconstruction operator.

Let us now assume that Q_i is conservative and let us prove that P_i is also conservative:

$$\begin{aligned}\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} P_i(x)\,dx &= \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \left(U_i^*(x) + Q_i(x)\right)\,dx \\ &= \overline{U}_i^* + V_i = \overline{U}_i^* + 0 = \overline{U}_i^*\end{aligned}$$

Finally, let P_i be the reconstruction operator that approximates a function $U(x)$ and let us show that it is high-order accurate provided that the stationary solutions of (1.0.1) are smooth:

$$U(x) - P_i(x) = U(x) - U_i^*(x) - Q_i(x) = O(\Delta x^p), \tag{2.1.3}$$

since Q_i is the reconstruction operator of order p applied to the fluctuations. □

2.2 Numerical integration

The initial cell averages of a initial condition U_0 are commonly approximated by quadrature formulas:

$$\overline{U}_i^0 = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U_0(x)\,dx \approx \widetilde{U}_i^0 = \sum_{m=1}^{M} \alpha^m U_0(x_i^m) = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U_0(x)\,dx + O(\Delta x^q), \tag{2.2.1}$$

with $q \geq p$ the order of the quadrature formula. Here, $\alpha^1, \ldots, \alpha^M$ and $x_i^1, \ldots, x_i^M$ denote, respectively, the weights and the nodes of the quadrature formula.

If the cell averages are obtained by applying quadrature formulas, exactly well-balanced reconstruction operators can still be defined by means of the following modification of the reconstruction procedure:

Algorithm 2.2.1. *Given a family of cell values $\{\widetilde{U}_i\}$, at every cell I_i:*

1. *Find, if possible, a stationary solution $U_i^*(x)$ in the cells belonging to the stencil of I_i, $\cup_{j\in\mathcal{S}_i} I_j$, such that:*
$$\sum_{m=1}^{M} \alpha^m U_i^*(x_i^m) = \widetilde{U}_i. \tag{2.2.2}$$
Otherwise, take $U_i^* \equiv 0$.

2. *Apply a standard reconstruction operator of order p to the cell values $\{V_j\}_{j\in S_i}$, called fluctuations, given by*
$$V_j = \widetilde{U}_j - \sum_{m=1}^{M} \alpha^m U_i^*(x_j^m), \quad j \in \mathcal{S}_i,$$
to obtain:
$$Q_i(x) = Q_i(x; \{V_j\}_{j\in\mathcal{S}_i}).$$

3. *Finally, define*
$$P_i(x) = U_i^*(x) + Q_i(x). \tag{2.2.3}$$

Moreover, the use of quadrature formulas to approximate the integral of the source term
$$\int_{x_{i-1/2}}^{x_{i+1/2}} S(P_i^t(x)) H_x \, dx \tag{2.2.4}$$
may destroy the well-balanced character of the method. This difficulty is avoided by rewriting the integral as proposed in [32]:
$$\begin{aligned}\int_{x_{i-1/2}}^{x_{i+1/2}} S(P_i^t(x)) H_x(x) \, dx = f\left(U_i^{t,*}(x_{i+1/2})\right) - f\left(U_i^{t,*}(x_{i-1/2})\right) \\ + \int_{x_{i-1/2}}^{x_{i+1/2}} \left((S(P_i^t(x)) - S(U_i^{t,*}(x))\right) H_x(x) \, dx,\end{aligned} \tag{2.2.5}$$
where $U_i^{t,*}$ denotes again the stationary solution found in the first step of the exactly well-balanced reconstruction procedure in Algorithm 2.2.1, at the cell I_i, at time t, and the reconstruction operator P_i^t is defined as (1.2.7).

Then, the integral of the source term is approximated as follows:
$$\begin{aligned}S_i^t = f\left(U_i^{t,*}(x_{i+1/2})\right) - f\left(U_i^{t,*}(x_{i-1/2})\right) \\ + \Delta x \sum_{m=1}^{M} \alpha^m \left(S(P_i^t(x_i^m)) - S(U_i^{t,*}(x_i^m))\right) H_x(x_i^m).\end{aligned} \tag{2.2.6}$$

The following Definition and Theorem represent the natural extensions of Theorem 2.0.1:

Definition 2.2.1. *The numerical method* (1.2.33) *is said to be exactly well-balanced for a stationary solution U^* of* (1.0.1) *if the vector of exact cell-averages of U^* (or their approximations if a quadrature formula is used) is an equilibrium of the ODE given by the numerical method* (1.2.33).

Theorem 2.2.1. *If the reconstruction operator P_i is exactly well-balanced for a stationary solution U^* of* (1.0.1)*, then the numerical method* (1.2.33)-(2.2.6) *is also exactly well-balanced for U^*, according to Definition 2.2.1.*

2.3 First- and second-order methods

In this thesis, the midpoint rule has been selected to approximate the integrals in the case of first- and second-order methods:

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} U(x)\,dx = U(x_i) + O(\Delta x^2). \tag{2.3.1}$$

Then, given an initial condition U_0, the sequence of initial cell averages $\{\widetilde{U}_i^0\}$ is given by

$$\widetilde{U}_i^0 = U_0(x_i),$$

that is, $\{\widetilde{U}_i^0\}$ is the family of point values of U_0.

Moreover, observe that, given a function U, the conservative property (1.2.4) for a reconstruction operator P_i of U reads as follows

$$P_i(x_i) = \widetilde{U}_i,$$

where $\widetilde{U}_i = U(x_i)$.

The algorithm of the well-balanced reconstruction procedure if the midpoint rule is applied is:

Algorithm 2.3.1. *Given a family of cell values $\{\widetilde{U}_i\}$, at every cell I_i:*

1. *Find, if possible, a stationary solution $U_i^*(x)$ defined in the cells belonging to the stencil of I_i, $\cup_{j\in\mathcal{S}_i} I_j$, such that:*

$$U_i^*(x_i) = \widetilde{U}_i, \tag{2.3.2}$$

i.e., look for $U_i^*(x)$ in the cells of the stencil of I_i which solves the Cauchy problem

$$\begin{cases} f(U)_x = S(U)H_x, \\ U(x_i) = \widetilde{U}_i. \end{cases} \tag{2.3.3}$$

Otherwise, take $U_i^* \equiv 0$.

2. *Apply a standard reconstruction operator of first-/second-order to the fluctuations given by*

$$V_j = \widetilde{U}_j - U_i^*(x_j), \quad j \in \mathcal{S}_i,$$

to obtain:

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

Observe that $V_i = 0$.

3. *Finally, define*

$$P_i(x) = U_i^*(x) + Q_i(x). \tag{2.3.4}$$

Remark 2.3.1. *In this work, the following choices are considered for the standard reconstruction operators:*

- *For the first-order methods, we consider the piecewise constant reconstruction operator as the standard one*

$$Q_i(x) = V_i = 0.$$

Therefore, the exactly well-balanced reconstruction operator is given by the solution of the Cauchy problem (2.3.3)*:*

$$P_i(x) = U_i^*(x).$$

- *For the second-order methods, a MUSCL reconstruction is considered:*

$$Q_i(x) = V_i + \Delta_i V(x - x_i) = \Delta_i V(x - x_i),$$

where $\Delta_i V$ *is the approximation of the spacial derivative obtained by using the mindmod or avg slope limiters described in Chapter 1. The second-order exactly well-balanced reconstruction operator writes as follows:*

$$P_i(x) = U_i^*(x) + \Delta_i V(x - x_i).$$

Taking everything into account, the numerical source term (2.2.6) is particularised for the first- and second-order schemes as follows:

$$S_i^t = f\left(U_i^{t,*}(x_{i+1/2})\right) - f\left(U_i^{t,*}(x_{i-1/2})\right) \tag{2.3.5}$$

since with the midpoint rule

$$\Delta x \left(S(P_i^t(x_i)) - S(U_i^{t,*}(x_i))\right) H_x(x_i) = 0. \tag{2.3.6}$$

2.4 Exactly well-balanced methods for a set of stationary solutions

The exactly well-balanced reconstruction procedure can be adapted to preserve not all the stationary solutions of (1.0.1), but a subset, denoted by $\mathcal{C}$. In this case, the first stage of the reconstruction procedure has to include a criterion to choose one of the elements of $\mathcal{C}$. In [32], two particular cases are detailed:

- Let us suppose that $\mathcal{C}$ is a k-parameter family of stationary solutions to preserve
$$U(x; C_1, \dots, C_k), \tag{2.4.1}$$
with $k < N$ and $C_1, \dots, C_k$ the k parameters. The nonlinear system to be solved in the first step is over-determined. We consider then a a subset of k indexes $\{j_1, \dots, j_k\}$ and the condition on the average in the first step is replaced by
$$\frac{1}{\Delta x}\int_{i-1/2}^{i+1/2} u^*_{i,j_s}(x; C_1, \dots, C_l)\, dx = \overline{u}_{i,j_s}, \quad s = 1, \dots, k, \tag{2.4.2}$$
or by
$$\sum_{m=1}^{M} \alpha^m u^*_{i,j_s}(x_i^m; C_1, \dots, C_l) = \widetilde{u}_{i,j_s}, \quad s = 1, \dots, k, \tag{2.4.3}$$
if a quadrature formula is used, where $u^*_{i,j}$, $\overline{u}_{i,j}$ and $\widetilde{u}_{i,j}$ denote, respectively, the j-th component of U^*_i, $\overline{U}_i$ and $\widetilde{U}_i$.
- If there is only one stationary solution $U^*(x)$ to preserve ($\mathcal{C}$ only has one element which is $U^*(x)$), the first step of the exactly well-balanced reconstruction procedure can be skipped by taking $U^*_i \equiv U^*$.

Note that, the exactly well-balanced reconstruction operators P_i so obtained are conservative in spite of the fact that the equalities (2.1.1) or (2.2.2) are not fulfilled. Moreover, observe that, in general, if the midpoint rule is considered for first- and second-order schemes, the fluctuation
$$V_i = \widetilde{U}_i - U^*_i(x_i)$$
does not vanish and the expression of reconstruction operator is

- For the first-order methods:
$$P_i(x) = U^*_i(x) + V_i = U^*_i(x) + \widetilde{U}_i - U^*_i(x_i); \tag{2.4.4}$$
- for the second-order methods:
$$P_i(x) = U^*_i(x) + V_i + \Delta_i V(x - x_i). \tag{2.4.5}$$

Therefore, the numerical source term for first- and second-order methods writes as follows:
$$S^t_i = f\left(U^{t,*}_i(x_{i+1/2})\right) - f\left(U^{t,*}_i(x_{i-1/2})\right) + \Delta x \left(S(P^t_i(x_i)) - S(U^{t,*}_i(x_i))\right) H_x(x_i). \tag{2.4.6}$$

Part II

Collection of manuscripts

High-order well-balanced methods for systems of balance laws: a control-based approach

This first paper, accepted for publication the 14^{th} November 2020 in *Applied Mathematics and Computation*, is focused on the design of high-order well-balanced finite volume methods for general systems of balance laws of the form (1.0.1), where H is a continuous function, using control techniques.

In Chapter 2, a general technique to design exactly well-balanced reconstruction operators based on the modification of standard ones such as MUSCL or CWENO was described. In that methodology, which consists of three stages, local nonlinear problems have to be solved at every cell and at every time step in the first one. In order to solve these problems, one has to look for a stationary solution whose cell average in the corresponding cell is given and to be able to extend it to the whole stencil, i.e., one has to look for the solution of problems of the form

$$\begin{cases} f(U)_x = S(U)H_x, \quad x \in I_i \\ \displaystyle\sum_{m=1}^{M} \alpha^m U(x_i^m) = W, \end{cases} \tag{2.4.7}$$

where W is a given state and α^m, x_i^m, $m = 1, \dots, M$ are, respectively, the weights and the points of the selected quadrature formula. Then, once the solution U_i^* has been found at the cell I_i, it has to be extended to all the cells of its stencil by solving two types of Cauchy problems:

$$\begin{cases} f(U)_x = S(U)H_x, \quad x \in I_j,\, j \in \mathcal{S}_i,\, j < i \\ U(x_{j+1/2}) = U_i^*(x_{j+1/2}), \end{cases} \tag{2.4.8}$$

and

$$\begin{cases} f(U)_x = S(U)H_x, \quad x \in I_j,\, j \in \mathcal{S}_i,\, j > i \\ U(x_{j-1/2}) = U_i^*(x_{j-1/2}). \end{cases} \tag{2.4.9}$$

Therefore, it is essential to know the expression of the solutions of (2.0.1) either in explicit or implicit form. There are problems, such as the shallow water equations, for which the solutions of (2.4.7), (2.4.8), (2.4.9) can be obtained by hand or by applying some standard iterative methods for nonlinear problems (see [79] and its references). This paper is focused in the case in which the analytical expression of the stationary solutions is not available or it is very expensive to compute: a strategy is proposed to approximate numerically the solutions of problems of the form (2.4.7), (2.4.8), (2.4.9). The reconstruction operators so obtained and, consequently, the numerical methods are not exactly well-balanced but well-balanced.

Problems of the form (2.4.8) and (2.4.9) are standard Cauchy problems and their solutions will be numerically approximated by using one-step numerical methods of the form

$$U_{m+1} = U_m + h\Phi_h(U_m), \quad m = 0, 1, \dots \tag{2.4.10}$$

forward or backward in space. More precisely, explicit RK methods are applied using a submesh of the stencil of step h that contains the intercells and the quadrature points.

Concerning problems of the form (2.4.7), we propose to reformulate the local nonlinear problems as control ones in which the control variable corresponds to the initial condition of a Cauchy problem, i.e. we look for the value at the left intercell that makes the solution to have the prescribed average. The local problems are written in functional form and the gradient of the functional is obtained on the basis of the adjoint variables. Newton's method is then applied to solve the control problems. The Cauchy problems related to the state and adjoint systems are numerically solved by using the selected explicit Runge-Kutta method. The main difficulty when applying these solvers comes from the backward integration, since the well-balanced property of the methods may be lost if the method is not symmetric in the following sense:

Definition 2.4.1. *The numerical one-step method* (2.4.10) *is said to be reversible or symmetric if it satisfies*

$$\Phi_h \circ \Phi_{-h} = Id, \quad \textit{or equivalently} \quad \Phi_h = \Phi_{-h}^{-1}. \tag{2.4.11}$$

However, there are not explicit symmetric methods (see [16]). Since our goal is to use explict RK methods in order not to increase the computational cost, we propose to take as control variable the value of the solution at the lefmost intercell of the stencil so that it is only necessary to apply the ODE solver forward in space.

Another difficulty comes from the fact that the ODE system satisfied by the stationary solution is not written in normal form, i.e.

$$U_x = G(x, U).$$

If sonic states are detected in the stencil, the Implicit Function Theorem cannot be applied. Consequently, the preservation of transonic stationary solutions (i.e., having at least one

sonic state and changing regime) becomes particularly difficult. In this paper, our proposal is to apply the standard reconstruction if a sonic point is detected in the stencil. In this case, the stationary solution is not preserved.

In the article, after presenting the method and algorithms, first-, second- and third-order well-balanced methods have been applied to the following PDEs:

- The Burgers equation with a nonlinear source term of the form

$$U_t + \left(\frac{U^2}{2}\right)_x = U^2. \tag{2.4.12}$$

  The choice of this simple problem is of major interest since the analytical expression of the stationary solutions is available, which allows us to compare the exactly well-balanced methods in [32] and well-balanced methods based on control techniques.

- The Burgers equation with sinusoidal source term of the form

$$U_t + \left(\frac{U^2}{2}\right)_x = \sin(U). \tag{2.4.13}$$

- The coupled Burgers equations with nonlinear source terms of the form

$$\begin{cases} (u_1)_t + \left(\dfrac{u_1^2}{2}\right)_x = 2u_1^2 + u_1u_2, \\ (u_2)_t + \left(\dfrac{u_2^2}{2}\right)_x = -u_1u_2 + 3u_2^2. \end{cases} \tag{2.4.14}$$

  This is an example of a more complex system of balance laws for which the expression of the stationary solutions is also available. Again, we have considered both exactly well-balanced and well-balanced methods.

- The shallow water equations, given by

$$\begin{cases} h_t + q_x = 0, \\ q_t + \left(\dfrac{q^2}{h} + \dfrac{gh^2}{2}\right)_x = ghH_x, \end{cases} \tag{2.4.15}$$

  where $q(x,t)$ is the discharge, $h(x,t)$ is the thickness, g is the gravity and $H(x)$ is the depth function measured from a fixed reference level.

- The compressible Euler equations with gravitational force, given by

$$\begin{cases} \rho_t + (\rho u)_x = 0, \\ (\rho u)_t + (\rho u^2 + p)_x = -\rho H_x, \\ (E)_t + (u(E+p))_x = -\rho u H_x. \end{cases} \tag{2.4.16}$$

Here, $\rho \geq 0$ is the density, u the velocity, $p \geq 0$ the pressure, E the total energy per unit volume, and $H(x)$ the gravitational potential. Moreover, the internal energy e is given by $\rho e = E - \frac{1}{2}\rho u^2$. Pressure is determined from e through the equation of state

$$p = (\gamma - 1)\rho e,$$

where $\gamma > 1$ is the adiabatic constant.

The variety of problems considered highlights the generality of the methods here introduced: it is enough to know f, S, H, G, ∇G to design well-balanced high-order methods for systems (1.0.1) with H a continuous function. The numerical tests confirm the accuracy and the well-balanced property of the methods. Moreover, they put on evidence that, in many cases, the extra computational cost due to the resolution of local problems and to the well-balanced modification of the reconstruction operator is lower than the one that would require to lead the discretization errors to (close to) zero machine by refining the mesh or increasing the order of non-well-balanced methods.

It has been also checked that the simple technique considered here to deal with transonic stationary solutions gives acceptable results.

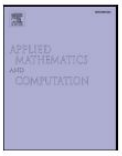
- **High-order well-balanced methods for systems of balance laws: a control-based approach.**
  I. Gómez-Bueno, M.J. Castro and C. Parés. Applied Mathematics and Computation 394 (2021): 125820. DOI: https://doi.org/10.1016/j.amc.2020.125820.

Applied Mathematics and Computation 394 (2021) 125820

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc



High-order well-balanced methods for systems of balance laws: a control-based approach

Irene Gómez-Bueno*, Manuel J. Castro, Carlos Parés

Dpto. Análisis Matemático, Estadística e Investigación Operativa y Matemática Aplicada, Universidad de Málaga Bulevar Louis Pasteur, 31, 29010 Málaga

ARTICLE INFO

Article history:
Received 13 November 2019
Revised 24 August 2020
Accepted 14 November 2020

Keywords:
Systems of balance laws
Well-balanced methods
Finite volume methods
High order methods
Reconstruction operators
Euler equations

ABSTRACT

In some previous works, two of the authors have introduced a strategy to develop high-order numerical methods for systems of balance laws that preserve all the stationary solutions of the system. The key ingredient of these methods is a well-balanced reconstruction operator. A strategy has been also introduced to modify any standard reconstruction operator like MUSCL, ENO, CWENO, etc. in order to be well-balanced. This strategy involves a non-linear problem at every cell at every time step that consists in finding the stationary solution whose average is the given cell value. So far this strategy has been only applied to systems whose stationary solution are known either in explicit or implicit form. The goal of this paper is to present a general implementation of this technique that can be applied to any system of balance laws. To do this, the nonlinear problems to be solved in the reconstruction procedure are interpreted as control problems: they consist in finding a solution of an ODE system whose average at the computation interval is given. These problems are written in functional form and the gradient of the functional is computed on the basis of the adjoint problem. Newton's method is applied then to solve the problems. Special care is put to analyze the effects of computing the averages and the source terms using quadrature formulas. To test their efficiency and well-balancedness, the methods are applied to a number of systems of balance laws, ranging from easy academic systems consisting of Burgers equation with some nonlinear source terms to the shallow water equations or Euler equations of gas dynamics with gravity effects.

Well-Balanced Reconstruction Operator for Systems of Balance Laws: Numerical Implementation

This manuscript, published the 26^{th} May 2021, is the third chapter of the book "Recent Advances in Numerical Methods for Hyperbolic PDE Systems", a volume of the SEMA SIMAI Springer Series. This volume corresponds to the Proceedings of the sixth edition of the International Conference "Numerical methods for hyperbolic problems" that took place in 2019 in Málaga (Spain). In the previous paper [1], a control-based approach has been considered in order to solve the local nonlinear problems in the first step of the well-balanced reconstruction procedure. Explicit RK methods have been considered to solve the state and adjoint systems and Newton's method is applied to solve the control problems.

Remember that the local problems in the well-balanced reconstruction procedure have to be solved at every cell and every time step. Then, an efficient numerical solver for the control problems is crucial. The choice of Newton's method was made after comparing the efficiency of several possibilities. These comparisons are reported in this book chapter. Together with Newton's method, descent-like methods with a suitable choice of the step were considered to solve the minimization problems

$$min_{U_0 \in R^N} J(U_0)$$

associated to the control problem, whose solution U_0 is the initial condition to be imposed at the leftmost point ot the stencil to compute the local stationary solution in a cell. These methods have the form

$$U_0^{k+1} = U_0^k - \rho_k d_k, \quad k = 0, 1, \ldots$$

where d_k is the descent direction and ρ_k, the step. In particular, we have considered gradient and conjugate gradient methods with a suitable stepsize. In order to choose the optimal step, different strategies have been compared:

- Two versions of a technique in which a varying step is obtained by multiplying or dividing the step in the previous iteration by a fixed parameter $\beta \in (0, 1)$ until some conditions related to the minimization of the cost function are fulfilled.

- A fixed stepsize.

- A strategy based on Armijo's rule, which is one of the inequalities of the Wolfe conditions. It requires two parameters: $\mu \in [0.01, 0.3]$ and $\beta \in [0.1, 0.8]$. Let us consider the function $\mathbf{J}(\rho) = J(U_0^k - \rho d_k)$. Then the first order approximation of $\mathbf{J}(\rho)$ at $\rho = 0$ is given by $\mathbf{J}(0) - \rho\,\mathbf{J}'(0)$. Define $\hat{\mathbf{J}}(\rho) = \mathbf{J}(0) - \mu\,\rho\,\mathbf{J}'(0)$. A stepsize ρ_k is considered acceptable by the Armijo Rule if $\mathbf{J}(\rho_k) \leq \hat{\mathbf{J}}(\rho_k)$. The stepsize ρ_{k+1} is chosen using the next rule: if $\mathbf{J}(\rho_k) > \hat{\mathbf{J}}(\rho_k)$, then $\rho_{k+1} = \beta\rho_k$, and otherwise take $\rho_{k+1} = \rho_k$.

- A technique based on the Wolfe conditions for unconstrained minimization problems, which are a set of inequalities for performing inexact line search, especially in quasi-Newton methods, first published by Philip Wolfe in 1969 [126]. It requires two parameters: $m_1 \in [0.01, 0.3]$ and $m_2 \in [0.5, 1]$. The stepsize is chosen using the next rule:

  a. Set $\rho_s^k = 0$ and $\rho_b^k = 0$.

  b. Compute the functions $\mathbf{J}(\rho_k)$ and $\hat{\mathbf{J}}(\rho_k)$.

  * If $\mathbf{J}(\rho_k) \leq \hat{\mathbf{J}}(\rho_k)$ and $\mathbf{J}'(\rho_k) \geq m_2\,\mathbf{J}'(0)$, take $\rho_{k+1} = \rho_k$ and stop the algorithm.
  * If $\mathbf{J}(\rho_k) > \hat{\mathbf{J}}(\rho_k)$, set $\rho_b^k = \rho_k$ and go to the step **c**.
  * If $\mathbf{J}(\rho_k) \leq \hat{\mathbf{J}}(\rho_k)$ and $\mathbf{J}'(\rho_k) < m_2\,\mathbf{J}'(0)$, set $\rho_s^k = \rho_k$ and go to **c**.

  c. Use the following rule to choose ρ_{k+1} :

  * If $\rho_b^k = 0$, take $\rho_{k+1} = a\rho_k$, where $a > 1$.
  * Else, take $\rho_{k+1} = \dfrac{\rho_s^k + \rho_b^k}{2}$.

The performance of the descent methods and Newton's method have been compared for a nonlinear scalar test control with state equation

$$U_x = \frac{\sin(U)}{U}. \tag{2.4.17}$$

This simple problem is solved 10000 times and the total number of iterations and the computational effort have been compared for the different options, concluding that Newton's method is more efficient.

The chapter ends with the application of the technique based on Newton's method to the design of well-balanced third order methods for two systems of balance laws:

- The scalar Burgers equation (2.4.13), which was also considered in [1].

- The shallow water system given by (2.4.15). Moreover, a modification of Newton's method has been considered in which the adjoint state is only recalculated every few iterations. The results put on evidence that the best choice is to compute the adjoint variables only once at the beginning of the algorithm.

The proposal in [1] to deal with critical points is also applied in this chapter: if a sonic point is detected in the stencil, the standard reconstruction in chosen.

As mentioned, the results show that the most efficient implementation to solve the local problems rewritten as control ones is to apply Newton's method. As concluded in [1], the well-balanced methods perform better than the standard ones, and in spite of the fact that an extra computational cost is required to build the well-balanced reconstruction operators, this computational effort is lower that the one necessary to reduce the discretization errors up to machine precision by considering higher order methods or by refining the meshes.

- **Well-Balanced Reconstruction Operator for Systems of Balance Laws: Numerical Implementation.**
  I. Gómez-Bueno, M.J. Castro and C. Parés. In: Muñoz-Ruiz, M.L., Parés, C., Russo, G. (eds) Recent Advances in Numerical Methods for Hyperbolic PDE Systems. SEMA SIMAI Springer Series, vol 28. Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-72850-2_3`.

Well-Balanced Reconstruction Operator for Systems of Balance Laws: Numerical Implementation

I. Gómez-Bueno, M. J. Castro, and C. Parés

Abstract In some previous works, two of the authors introduced a strategy to develop high-order well-balanced numerical methods for 1d systems of balance laws. There, a strategy which allows us to modify any standard reconstruction operator in order to be well-balanced was also described. This strategy involves a nonlinear problem at every cell, at every time step, that consists in finding the stationary solution whose average is the given cell value. Our goal is to present a general efficient implementation that can be applied to any system of balance laws by interpreting these nonlinear problems as control problems that are rewritten in functional form. Newton's and descent methods are applied and compared. Applications to the Burgers' equation with a nonlinear source term and to the 1d shallow water model are finally shown.

I. Gómez-Bueno (✉) · M. J. Castro · C. Parés
University of Málaga, Málaga, Spain
e-mail: igomezbueno@uma.es

M. J. Castro
e-mail: mjcastro@uma.es

C. Parés
e-mail: pares@uma.es

M. L. Muñoz-Ruiz et al. (eds.), *Recent Advances in Numerical Methods for Hyperbolic PDE Systems*,
SEMA SIMAI Springer Series 28, https://doi.org/10.1007/978-3-030-72850-2_3

Collocation Methods for High-Order Well-Balanced Methods for Systems of Balance Laws

This paper, accepted for publication the 25^{th} July 2021 in *Mathematics*, is also focused on the design of high-order well-balanced finite volume methods for general systems of balance laws of the form (1.0.1), where H is a continuous function. In this work, we propose to solve the local nonlinear problems arising in the first step of the well-balanced reconstruction procedure by applying collocation RK methods (see [81]). These implicit RK methods can be introduced in two different ways:

- On the one hand, they can be interpreted as standard Runge-Kutta methods with an associated Butcher tableau of the form

$$\begin{array}{c|ccc} c_1 & a_{1,1} & \dots & a_{1,s} \\ c_2 & a_{2,1} & \dots & a_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s,1} & \dots & a_{s,s} \\ \hline & b_1 & \dots & b_s. \end{array}$$

  Let us then consider an index i_0, and let us show how these methods work when applied to a standard Cauchy problem of the form

$$\begin{cases} U_x = G(U, x), \\ U(x_{i_0-1/2}) = U^{i_0-1/2}, \end{cases} \tag{2.4.18}$$

  in a uniform mesh of nodes $x_{i+1/2} = x_{i-1/2} + h$, $i = i_0, i_0+1, \dots$ The approximations obtained with the collocation RK methods at the nodes are computed by:

$$U^{i+1/2} = U^{i-1/2} + h\,\Phi_h(U^{i-1/2}), \quad i = i_0, i_0+1, \dots \tag{2.4.19}$$

  with

$$\Phi_h(U^{i-1/2}) = \sum_{j=1}^{s} b_j K_i^j,$$

where $K_i^1, \ldots, K_i^s$ fulfil the nonlinear system

$$K_i^m = G\left(U^{i-1/2} + h\sum_{l=1}^{s} a_{m,l}K_i^l, x_i^m\right), \quad m = 1, \ldots, s, \tag{2.4.20}$$

where

$$x_i^m = x_{i-1/2} + c_m h, \quad m = 1, \ldots, s. \tag{2.4.21}$$

In this work, the $2s$-th order Gauss-Legendre methods have been considered with uniform step $h = \Delta x$, since in this case $b_1, \ldots, b_s$ and $x_i^1, \ldots, x_i^s$ are, respectively, the weights and the quadrature nodes of the Gauss quadrature formula in the interval $[x_{i-1/2}, x_{i+1/2}]$, so

$$s = M \text{ and } b_m = \alpha^m,\ m = 1, \ldots, M.$$

- On the other hand, the approximation obtained at $x_{i+1/2}$ can be interpreted as

$$U^{i+1/2} = P_i(x_{i+1/2}),$$

where P_i is the only polynomial of degree s satisfying

$$\begin{cases} P_i'(x_i^m) = G(P_i(x_i^m), x_i^m), & m = 1, \ldots, s, \\ P_i(x_{i-1/2}) = U^{i-1/2}. \end{cases} \tag{2.4.22}$$

Because of this double interpretation, it can be easily shown that the collocation RK methods are symmetric in the sense of the Definition 2.4.1 and, consequently, losses of the well-balanced property related to the backward integration in space will not appear.

In addition, this work provides the theoretical basis for the design of well-balanced high-order numerical methods: Definitions 2.0.1 and 2.2.1, Theorems 2.1.1 and 2.2.1 and Algorithm 2.2.1 that constitute the theoretical framework for exactly well-balanced methods are here extended to well-balanced methods.

Again, since the ODEs are not in normal form, special care has to be taken in resonant problems. Unlike previous works, in this paper a general technique with implicit methods to deal with resonant problems for any system of balance laws has been developed. Solving the local nonlinear problems requires the resolution of linear systems of the form

$$J_f(W)K = S(W)H_x(\bar{x}). \tag{2.4.23}$$

where W is a given state and $\bar{x}$ a given point. Let us suppose that W is a sonic state and our problem is, consequently, resonant. In this case, (2.4.23) may have not solution or may have more than one (indeed, infinitely many):

- In the former case, W cannot be the value at $\bar{x}$ of any stationary solution. If this situation is detected during the well-balanced reconstruction procedure, it is stopped and the standard reconstruction is applied.

- In the latter one, since there are infinitely many solutions, the notion of *admissible* solution is introduced: we say that a solution K of (2.4.23) is admissible if there exists a C^1 stationary solution U^* such that

$$U^*(\bar{x}) = W, \quad U^*_x(\bar{x}) = K.$$

  If K is an admissible solution, then one has:

$$\lim_{x\to\bar{x}} J_f(U^*)^{-1} S(U^*) H_x(x) = K, \tag{2.4.24}$$

  If this limit can be computed analytically, this value is then chosen as the solution of (2.4.23).

In the paper, the shallow water system with topography (2.4.15) is considered in order to illustrate this procedure. It can be easily checked that system (2.4.23), which in the case of the shallow water system becomes

$$\begin{bmatrix} 0 & 1 \\ -u^2+gh & 2u \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} 0 \\ ghH_x(x) \end{bmatrix}, \tag{2.4.25}$$

reduces to:

$$\begin{bmatrix} 0 & 1 \\ 0 & 2u \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} 0 \\ ghH_x(x) \end{bmatrix}, \tag{2.4.26}$$

when U^* is a critical state, i.e. when $u^2 = gh$. Therefore, the system has solutions only if $H_x(x) = 0$: in this case the solutions are

$$K = \alpha[1, 0]^T, \quad \alpha \in \mathbb{R}.$$

In fact, a smooth stationary solution can only reach a critical state at a minimum point x^c of the depth function H (see [79]). The stationary solutions are given in implicit form by:

$$q = C_1, \quad \frac{q^2}{2h^2} + gh - gH = C_2, \quad C_1, C_2 \in \mathbb{R}, \tag{2.4.27}$$

If we now differentiate the second equation of (2.4.27) with respect to x, and we use the condition $H_x(x_c) = 0$, at $x = x_c$ one has

$$\frac{qq_x}{h^2} + h_x\left(g - \frac{q^2}{h^3}\right) = 0$$

which, assuming $h_x(x_c) \neq 0$, implies

$$h^c(q) = \frac{q^{2/3}}{g^{1/3}}. \tag{2.4.28}$$

Let us then compute the limit

$$\lim_{x\to x^c} \frac{ghH_x}{-u^2+gh},$$

that can be rewritten as follows:

$$\lim_{x\to x^c} \frac{gH_x}{g-\dfrac{q^2}{h^3}}, \tag{2.4.29}$$

to determine the value of the derivative h_x at the critical point x^c. Observe that L'Hôpital's rule can be applied since it is a $0/0$ indeterminate limit. Some easy computation leads to

$$h_x(x^c) = \pm\sqrt{\frac{q^{2/3}H_{xx}(x^c)}{3g^{1/3}}}. \tag{2.4.30}$$

The above expression shows that if $q \neq 0$ and $H_{xx}(x_c) \neq 0$, it is $h_x(x_c) \neq 0$, thus justifying the assumption used before. Then, the chosen solution of (2.4.23) will be

$$K = \left[\pm\sqrt{\frac{q^{2/3}H_{xx}(x^c)}{3g^{1/3}}}, 0\right]^T.$$

Let us now illustrate this technique in the particular case of the Euler equations with gravity, given by

$$\begin{cases} \rho_t + (\rho u)_x = 0, \\ (\rho u)_t + (\rho u^2 + p)_x = -\rho H_x, \\ (E)_t + (u(E+p))_x = -\rho u H_x, \end{cases} \tag{2.4.31}$$

where $\rho \geq 0$ is the density, u the velocity, $p \geq 0$ the pressure, E the total energy per unit volume, and $H(x)$ the gravitational potential. Moreover, the internal energy e is given by $\rho e = E - \dfrac{1}{2}\rho u^2$. Pressure is determined from e through the equation of state

$$p = (\gamma - 1)\rho e, \tag{2.4.32}$$

where $\gamma > 1$ is the adiabatic constant.

The eigenvalues of the system are $\lambda_1 = u - c$, $\lambda_2 = u$ and $\lambda_3 = u + c$, where where $c = \sqrt{\gamma p/\rho}$ is the sound speed. The Mach number defined by

$$M(U) = \frac{|u|}{c}, \tag{2.4.33}$$

characterizes the flow regime: subsonic ($M < 1$), sonic ($M = 1$) or supersonic ($M > 1$).

Supposing that the system is strictly hyperbolic and considering (2.4.32), the stationary solutions satisfy the ODE system:

$$\begin{cases} q_x = 0, \\ \dfrac{d\hat{U}}{dx} = G(x, \hat{U}), \end{cases} \tag{2.4.34}$$

where

$$\hat{U} = \begin{pmatrix} \rho \\ E \end{pmatrix}, \quad G(x, \hat{U}) = - \begin{pmatrix} \dfrac{\rho}{c^2 - u^2} \\ \dfrac{\rho}{\gamma - 1} \left(1 + \dfrac{3 - \gamma}{2} \dfrac{u^2}{c^2 - u^2} \right) \end{pmatrix} H_x.$$

Notice that for regular solutions of the Euler equations the entropy is constant along material lines. So that, for stationary solutions, $us_x = 0$, where s denote the entropy density. Therefore if u does not vanish it follows that these stationary solutions are isentropic.

One-dimensional steady adiabatic flow are given in implicit form by:

$$q = C_1, \quad \frac{q^2}{2\rho^2} + h + H = C_2, \quad p\rho^{-\gamma} = C_3 \quad C_1, C_2, C_3 \in \mathbb{R}, \tag{2.4.35}$$

where

$$h = e + \frac{p}{\rho}. \tag{2.4.36}$$

Since the momentum is constant for any stationary solution, system (2.4.23) will be only solved for the density and energy. Therefore, system (2.4.23) becomes in our case

$$\begin{bmatrix} \frac{1}{2}(\gamma - 3)u^2 & \gamma - 1 \\ \frac{\gamma - 2}{2}u^3 - \frac{1}{\gamma - 1}uc^2 & \gamma u \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} -\rho H_x \\ -\rho u H_x \end{bmatrix}. \tag{2.4.37}$$

Let us suppose that U^* is a critical state, i.e. $u^2 = c$. Then (2.4.37) reduces to

$$\begin{bmatrix} \frac{1}{2}(\gamma - 3)u^2 & \gamma - 1 \\ \frac{\gamma(\gamma - 3)}{2(\gamma - 1)}u^3 & \gamma u \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} -\rho H_x \\ -\rho u H_x \end{bmatrix}. \tag{2.4.38}$$

Therefore, system (2.4.23) has solutions only if $H_x(x) = 0$: in this case, given K_1, the value K_2 is computed as follows

$$K_2 = \frac{\gamma - 3}{2(\gamma - 1)} u^2 K_1.$$

Differentiating the second relation of (2.4.35) with respect to x, and using the relation $H_x(x_c) = 0$ and the first relation of (2.4.35), one obtains, at $x = x_c$,

$$- \left((\gamma - 1)\frac{u^2}{\rho} - \gamma \frac{E}{\rho^2} \right) \rho_x + \gamma \frac{E}{\rho} E_x = 0. \tag{2.4.39}$$

Analogously, differentiating the third relation of (2.4.35) with respect to x, one has

$$E_x = \left(\gamma\frac{E}{\rho} - (\gamma+1)\frac{q^2}{2\rho^2}\right)\rho_x. \tag{2.4.40}$$

Substituting (2.4.40) in (2.4.39), one obtains

$$-\left(\left(-1-\frac{\gamma(\gamma-1)}{2}\right)\frac{u^2}{\rho} - \gamma(\gamma-1)\frac{E}{\rho^2}\right)\rho_x = 0, \tag{2.4.41}$$

and some computations lead to

$$\rho^c(q) = \left(\frac{q^2}{\gamma C_3}\right)^{\frac{1}{1+\gamma}}. \tag{2.4.42}$$

With this information in mind, let us compute the limit

$$\lim_{x\to x^c} -\frac{H_x}{-\frac{q^2}{\rho^3} + \gamma C_3 \rho^{\gamma-2}},$$

to determine the value of h_x at x^c. This is a $0/0$ indeterminate limit and L'Hôpital's rule can be applied again. Some easy computation leads to

$$\rho_x(x^c) = \pm\sqrt{\frac{-H_{xx}(x^c)}{3q^2\left(\frac{q^2}{\gamma C_3}\right)^{\frac{-4}{1+\gamma}} + \gamma(\gamma-2)C_3\left(\frac{q^2}{\gamma C_3}\right)^{\frac{\gamma-3}{1+\gamma}}}}. \tag{2.4.43}$$

Then, the chosen solution of (2.4.23) will be

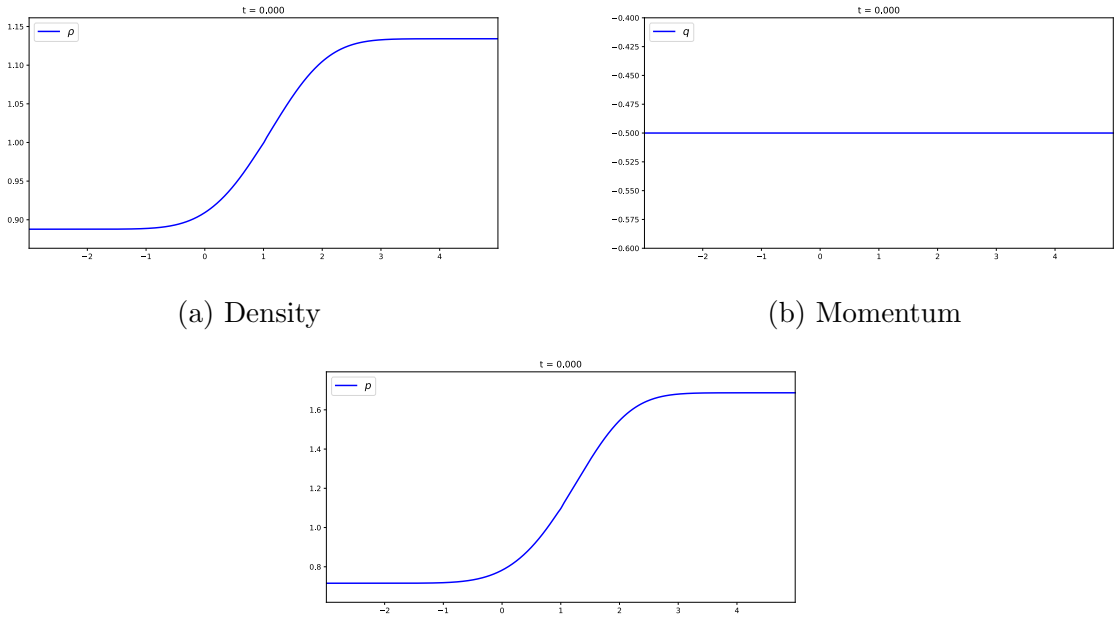
$$K_1 = \pm\sqrt{\frac{-H_{xx}(x^c)}{3q^2\left(\frac{q^2}{\gamma C_3}\right)^{\frac{-4}{1+\gamma}} + \gamma(\gamma-2)C_3\left(\frac{q^2}{\gamma C_3}\right)^{\frac{\gamma-3}{1+\gamma}}}}, \quad K_2 = \frac{\gamma-3}{2(\gamma-1)}u^2K_1. \tag{2.4.44}$$

In order to check the previous results, two numerical tests with a first-order well-balanced scheme have been performed. The first experiment is intended to check the correct behaviour of the method in presence of sonic states. We consider here a smooth transonic stationary solution for the Euler equations with gravity in the interval $[-3, 5]$. Here, $\gamma = 5/3$. The integration time is $t = 150$ and H is given by

$$H(x) = \frac{1}{200}e^{-(x-1)^2}. \tag{2.4.45}$$

The initial condition (see Figure (2.1) corresponds to the transonic stationary solution U^* with a sonic state at $x^c = 1$ corresponding to the choices

$$q^c = -\frac{1}{2}, \quad \rho^c = 1, \quad p\rho^{-\gamma} = \frac{1}{4\gamma}. \tag{2.4.46}$$



(a) Density

(b) Momentum

(c) Pressure

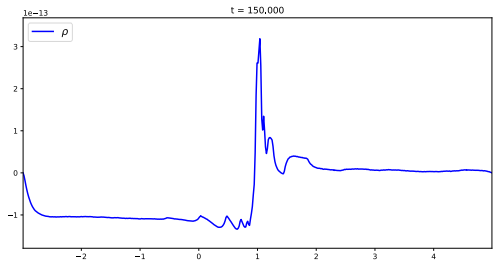
Figure 2.1: Euler equations with gravity. Initial condition: smooth transonic stationary solution.

	ρ	q	E
Error	4.98e-13	6.22e-15	1.23e-13

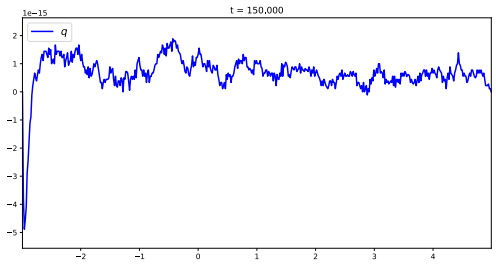
Table 2.1: Euler equations with gravity: smooth transonic stationary solution. L^1-errors at time $t = 150$ for the 500-cell mesh.

The difference between the stationary and the numerical solution computed with a mesh with 500 cells is shown in Figure 2.2. L^1-errors are shown in Table 2.1. Notice that the stationary solution is preserved with machine precision.

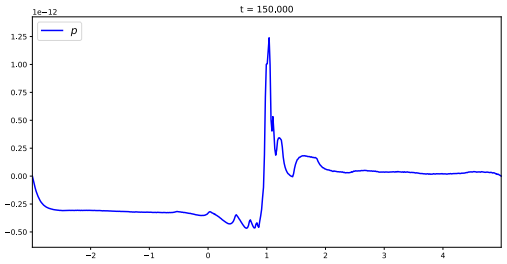
In the second test case, we consider the evolution of a small perturbation of the previous transonic stationary solution. More precisely, a small perturbation of the density of amplitude $\Delta\rho = 10^{-4}$ is set at the interval $[4.0, 4.1]$, that is in the subsonic region. Therefore, this perturbation splits into three waves, two moving downstream and the other upstream. Figure 2.3 shows the difference between the stationary and the numerical solution at time $t = 5.5$ for a 500-cell mesh. Let us point out that the solution presents a small spurious oscillation in the location of the critical point for density and pressure. This spurious oscillation decreases with respect to time and Table 2.2 show the L^1-errors at time $t = 150$.



(a) Density



(b) Momentum



(c) Pressure

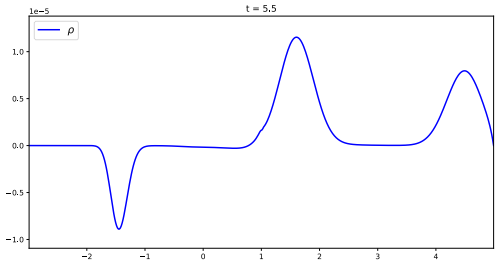
Figure 2.2: Euler equations with gravity: smooth transonic stationary solution. Differences between the stationary and the numerical solution given by a first-order well-balanced method at time $t = 150$ for the 500-cell mesh.

	ρ	q	E
Error	2.73e-11	5.94e-13	7.86e-12

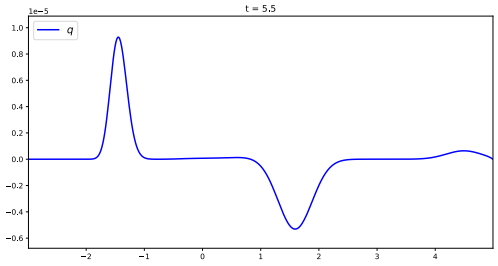
Table 2.2: Euler equations with gravity: perturbation of a smooth transonic stationary solution. L^1-errors at time $t = 150$ for the 500-cell mesh.

In the paper, first-, second- and third-order well-balanced methods designed within this framework have been applied to several systems of balance laws, ranging from simple scalar problems to more complex ones. The following PDEs have been considered:

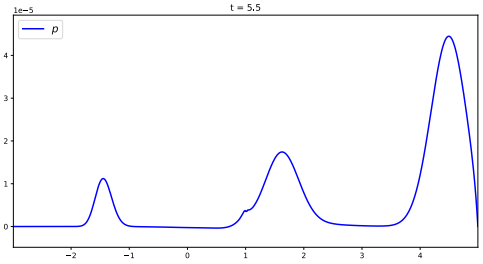
- Following [1], we start with the Burgers equation with a nonlinear source term of the form (2.4.12), whose stationary solutions are available. Not only standard and exactly well-balanced methods have been compared, but also well-balanced methods based on control techniques vs. the application of collocation RK methods.
- The Burgers equation with sinusoidal source term (2.4.13) that was also considered in the previous works.
- The shallow water equations (2.4.15).



(a) Density



(b) Momentum



(c) Pressure

Figure 2.3: Euler equations with gravity: perturbation of a smooth transonic stationary solution. Differences between the stationary and the numerical solution given by a first-order well-balanced method at time $t = 5.5$ for the 500-cell mesh.

- The shallow water model with Manning friction, given by

$$\begin{cases} h_t + q_x = 0, \\ q_t + \left(\dfrac{q^2}{h} + \dfrac{1}{2}gh^2\right)_x = ghH_x - \dfrac{kq|q|}{h^\eta}, \end{cases} \tag{2.4.47}$$

  where k is the Manning friction coefficient and η is set to $\frac{7}{3}$.

- The compressible Euler equations with gravitational force.

The large number of performed tests show that the strategy based on the application of collocation RK methods is more efficient than the control-based approach. In addition, the experiments show that the general strategy introduced in the paper to deal with resonant problems improves the numerical results obtained for transcritical stationary solutions compared to the simple technique used in previous works, in which the standard reconstruction operators were used when a sonic state was detected.

- **Collocation Methods for High-Order Well-Balanced Methods for Systems of Balance Laws.**
  I. Gómez-Bueno, M.J. Castro, C. Parés and G. Russo. Mathematics 9.15 (2021): 1799. DOI: `https://doi.org/10.3390/math9151799`.

mathematics

MDPI

Article

Collocation Methods for High-Order Well-Balanced Methods for Systems of Balance Laws

Irene Gómez-Bueno [1,*], Manuel Jesús Castro Díaz [1], Carlos Parés [1] and Giovanni Russo [2]

[1] Departamento de Análisis, Estadística e I.O. y Matemática Aplicada, University of Málaga, Avda. Cervantes, 2, 29071 Málaga, Spain; mjcastro@uma.es (M.J.C.D.); pares@uma.es (C.P.)
[2] Dipartimento di Matematica ed Informatica, University of Catania, Viale Andrea Doria, 6, 95125 Catania, Italy; russo@dmi.unict.it
* Correspondence: igomezbueno@uma.es

Abstract: In some previous works, two of the authors introduced a technique to design high-order numerical methods for one-dimensional balance laws that preserve all their stationary solutions. The basis of these methods is a well-balanced reconstruction operator. Moreover, they introduced a procedure to modify any standard reconstruction operator, like MUSCL, ENO, CWENO, etc., in order to be well-balanced. This strategy involves a non-linear problem at every cell at every time step that consists in finding the stationary solution whose average is the given cell value. In a recent paper, a fully well-balanced method is presented where the non-linear problems to be solved in the reconstruction procedure are interpreted as control problems. The goal of this paper is to introduce a new technique to solve these local non-linear problems based on the application of the collocation RK methods. Special care is put to analyze the effects of computing the averages and the source terms using quadrature formulas. A general technique which allows us to deal with resonant problems is also introduced. To check the efficiency of the methods and their well-balance property, they have been applied to a number of tests, ranging from easy academic systems of balance laws consisting of Burgers equation with some non-linear source terms to the shallow water equations—without and with Manning friction—or Euler equations of gas dynamics with gravity effects.

Keywords: systems of balance laws; well-balanced methods; finite volume methods; high order methods; reconstruction operators; collocation methods; shallow water equations; Euler equations

Citation: Gómez-Bueno, I.; Castro Díaz, M.J.; Parés, C.; Russo, G. Collocation Methods for High-Order Well-Balanced Methods for Systems of Balance Laws. *Mathematics* **2021**, *9*, 1799. https://doi.org/10.3390/math9151799

Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws

This last paper [4], accepted for publication the 24^{th} September 2022 in *Applied Numerical Mathematics*, is focused on the design of implicit and semi-implicit high-order well-balanced finite volume methods for general systems of balance laws of the form (1.0.1), where H is a continuous function.

In the previous works, we have obtained semi-discrete numerical methods for (1.0.1) of the form

$$\begin{aligned}\frac{d\widetilde{U}_i(t)}{dt} = &-\frac{1}{\Delta x}\left(F^t_{i+1/2} - F^t_{i-1/2}\right) + \frac{1}{\Delta x}\left(f\left(U_i^{t,*,i+1/2}\right) - f\left(U_i^{t,*,i-1/2}\right)\right)\\ &+ \sum_{m=1}^{M} \alpha^m \left(S(P_i^{t,m}) - S(U_{i,i}^{t,*,m})\right) H_x(x_i^m),\end{aligned} \tag{2.4.48}$$

where

$$U_i^{t,*,i\pm1/2}, \quad U_{i,i}^{t,*,m},\ m = 1,\dots,M, \tag{2.4.49}$$

are the approximations at the intercells and quadrature points of the stationary solution that solves the local problem in the first step of the well-balanced reconstruction procedure at the cell I_i at time t; $P_i^{t,m}$, $m = 1,\dots,M$ represent the value at the quadrature points of the well-balanced reconstructions obtained from the sequence $\{\widetilde{U}_i(t)\}$; and

$$F^t_{i+1/2} = \mathbb{F}(U^{t,-}_{i+1/2}, U^{t,+}_{i+1/2}), \tag{2.4.50}$$

where

$$U^{t,-}_{i+1/2} = P_i^{t,i+1/2}, \quad U^{t,+}_{i-1/2} = P_i^{t,i-1/2}. \tag{2.4.51}$$

Here $P_i^{t,i\pm1/2}$ represent the value at the intercells of the well-balanced reconstructions obtained from the vector $\{\widetilde{U}_i(t)\}$.

If (1.0.1) contains stiff terms, implicit and semi-implicit schemes could be obtained by applying implicit or semi-implicit RK methods to the ODE (2.4.48). However, this

strategy may lead to too complex nonlinear algebraic systems that may be unfeasible due to the high-order well-balanced reconstructions at every stage, which involves the search for local steady states related to the unknown solution at time t^{n+1}. In order to avoid this, once the cell averages of the solution at time t^n, $\{\widetilde{U}_i^n\}$, have been computed, where

$$\widetilde{U}_i^n = \widetilde{U}_i(t^n),$$

the cell averages in the time interval $t \in [t^n, t^{n+1}]$ are written as follows:

$$\widetilde{U}_i(t) = \widetilde{U}_i^n + U_i^f(t), \quad \forall i, \tag{2.4.52}$$

where the time fluctuations $U_i^f(t)$ are the solutions, at every cell I_i, of the ODE:

$$\begin{aligned}\frac{dU_i^f(t)}{dt} = &-\frac{1}{\Delta x}\left(F_{i+1/2}^t - F_{i-1/2}^t\right) + \frac{1}{\Delta x}\left(f\left(U_i^{n,*,i+1/2}\right) - f\left(U_i^{n,*,i-1/2}\right)\right) \\ &+ \sum_{m=1}^{M} \alpha^m \left(S(P_i^{t,m}) - S(U_{i,i}^{n,*,m})\right) H_x(x_i^m),\end{aligned} \tag{2.4.53}$$

with initial condition

$$U_i^f(t^n) = 0. \tag{2.4.54}$$

If this ODE system were to be solved exactly one would have then

$$\widetilde{U}_i^{n+1} = \widetilde{U}_i^n + U_i^f(t^{n+1}). \tag{2.4.55}$$

Instead, what we propose is to consider different reconstructions in (2.4.53):

$$P_i^{t,m} = P_i^{n,m} + \widetilde{Q}_i^t(x_i^m), \quad m = 1, \ldots, M, \tag{2.4.56}$$

where

$$P_i^{n,m}, \quad m = 1, \ldots, M, \tag{2.4.57}$$

is the discrete well-balanced reconstruction operator obtained from the sequence $\{\widetilde{U}_i^n\}$. Finally, $\widetilde{Q}_i^t$ is an 'easy' standard reconstruction operator defined in terms of the time fluctuations $\{U_j^f(t)\}_{j\in\widetilde{\mathcal{S}}_i}$,

$$\widetilde{Q}_i^t(x) = \widetilde{Q}_i^t(x; \{U_j^f(t)\}_{j\in\widetilde{\mathcal{S}}_i}). \tag{2.4.58}$$

Therefore, the discrete reconstruction operator considered in (2.4.53) is the sum of the well-balanced reconstruction operator at time t^n plus a standard reconstruction operator of the time fluctuations. Notice that this reconstruction operator is not the same as the one described in the previous works: in order to reduce the computational cost and complexity of the systems, $\widetilde{Q}_i^t$ uses the information given by the approximated solution at time t^n to compute the necessary coefficients and smoothness indicators. Observe that, with this modification of (2.4.53), the well-balanced reconstruction is only computed once per time iteration.

In the paper, the two following situations are considered:

1. both the numerical flux and the source are stiff;

2. some terms of (1.0.1) are stiff, but not all of them.

The first case requires an implicit treatment of both the source and the flux terms. In such case an implicit RK method is applied to solve (2.4.53). Following Section 1.2.5, a stiffly accurate DIRK method is applied to numerically solve the Cauchy problem. Then, computation of the time fluctuations reads as follows:

$$\begin{aligned} U_i^{f,k} &= \sum_{l=1}^{k-1} \gamma_{k,l} U_i^{f,l} + \gamma \Delta t L_i^k, \quad k = 1, \dots, s, \\ U_i^{f,n+1} &= U_i^{f,s}, \end{aligned}$$

for some coefficients $\gamma_{k,l}$, where

$$\begin{aligned} L_i^k = &-\frac{1}{\Delta x}\left(F_{i+1/2}^k - F_{i-1/2}^k\right) + \frac{1}{\Delta x}\left(f\left(U_i^{n,*,i+1/2}\right) - f\left(U_i^{n,*,i-1/2}\right)\right) \\ &+ \sum_{m=1}^{M} \alpha^m \left(S(P_i^{k,m}) - S(U_{i,i}^{n,*,m})\right) H_x(x_i^m),\ k = 1, \dots, s. \end{aligned} \tag{2.4.59}$$

In the second case, the problem is rewritten as follows:

$$U_t + f^1(U)_x + f^2(U)_x = S^1(U)H_x + S^2(U), \tag{2.4.60}$$

with f^1 and S^1 non stiff, and f^2 and S^2 stiff. An IMEX methods with double Butcher tableau is applied to (2.4.53) (see Section 1.2.5) , in which the non stiff terms are treated explicitly, while the stiff terms are treated implicitly. We select numerical fluxes $F^i(u_l, u_r)$, $i = 1, 2$ consistent with f^i, $i = 1, 2$, and the the method for the time fluctuations writes as follows:

$$\begin{aligned} U_i^{f,k} &= \Delta t \sum_{l=1}^{k-1} \tilde{a}_{k,l} L_i^{1,l} + \Delta t \sum_{l=1}^{k-1} a_{k,l} L_i^{2,l} + \Delta t \gamma L_i^{2,k}, \quad k = 1, \dots, s, \\ U_i^{f,n+1} &= \Delta t \sum_{l=1}^{s} \tilde{b}_i L_i^{1,l} + \Delta t \sum_{l=1}^{s-1} a_{s,i} L_i^{2,l} + \Delta t \gamma L_i^{2,s}, \end{aligned}$$

where

$$\begin{aligned} L_i^{1,k} = &-\frac{1}{\Delta x}\left(F_{i+1/2}^{1,k} - F_{i-1/2}^{1,k}\right) + \frac{1}{\Delta x}\left(f^1\left(U_i^{n,*,i+1/2}\right) - f^1\left(U_i^{n,*,i-1/2}\right)\right) \\ &+ \sum_{m=1}^{M} \alpha^m \left(S^1(P_i^{k,m}) - S^1(U_{i,i}^{n,*,m})\right) H_x(x_i^m); \\ L_i^{2,k} = &-\frac{1}{\Delta x}\left(F_{i+1/2}^{2,k} - F_{i-1/2}^{2,k}\right) + \frac{1}{\Delta x}\left(f^2\left(U_i^{n,*,i+1/2}\right) - f^2\left(U_i^{n,*,i-1/2}\right)\right) \\ &+ \sum_{m=1}^{M} \alpha^m \left(S^2(P_i^{k,m}) - S^2(U_{i,i}^{n,*,m})\right); \end{aligned} \tag{2.4.61}$$

for $k = 1, \ldots, s$.

In the paper, the well-balanced property of the fully discrete methods is proved. Despite of the generality of the methodology introduced here only first- and second-order methods have been implemented in this paper. In order to check the technique, it has been applied to several balance laws:

- The linear transport equation with source term:

$$U_t + (cU)_x = \alpha U, \ c, \alpha \in \mathbb{R} \setminus \{0\}. \tag{2.4.62}$$

- The Burgers equation with a nonlinear source term (2.4.12).

- The shallow water equations without friction (2.4.15).

- The shallow water equations with Manning friction (2.4.47).

The well-balanced property of the methods has been checked and also the ability of the methods to recover a stationary solution when a small perturbation of that stationary solution is set as initial condition. Moreover, order tests are performed for different systems to check the accuracy of the methods. Finally, the shock-capturing property of the methods and the ability of convergence to steady states for long time periods when the initial condition is far from equilibrium has been also studied.

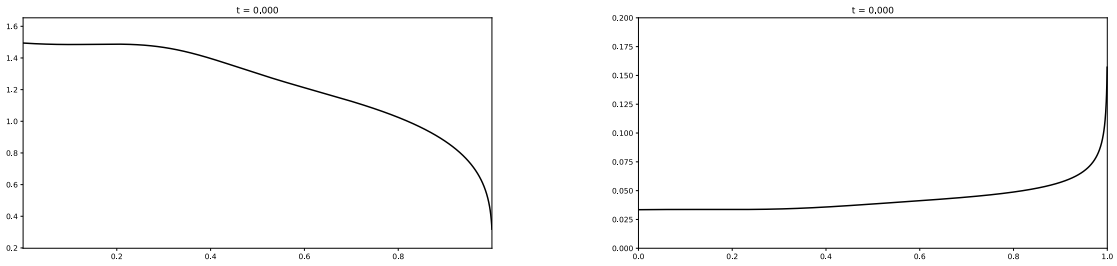
The numerical experiments for the shallow water equations with Manning friction included in the paper consider small values of the Manning friction coefficient k. Let us now perform a new test considering a large value of k to check what happens in the large friction limit. We consider $k = 5000$ and the space interval is $[0, 1]$. The final time is $t = 0.1$ and the depth function is constant $H(x) = 1$. The initial condition $U_0(x) = [h_0(x), q_0(x)]^T$ is given by

$$h_0(x) = h^*(x) + 0.1e^{-25(x-3)^2}, \quad q_0(x) = q^*(x),$$

where $U^*(x) = [h^*(x), q^*(x)]^T$ is the stationary solution corresponding to

$$\begin{cases} (-u^2 + gh)\, h_x = ghH_x - \dfrac{kq|q|}{h^\eta}, \\ q_x = 0, \end{cases} \tag{2.4.63}$$

with final conditions $h(1) = 0.3$ and $q(1) = 0.05$ (see Figure (2.4)).



(a) h (b) u

Figure 2.4: Shallow water with large value of the Manning friction coefficient. Initial condition: perturbation of a stationary solution.

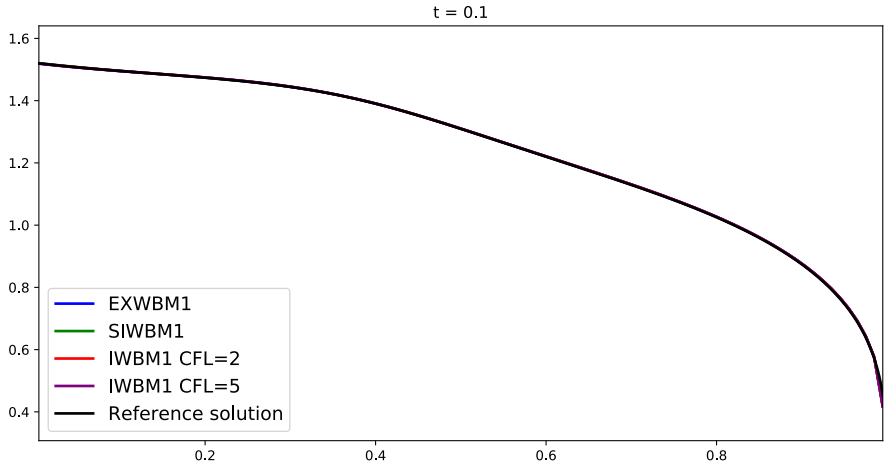
The first-order explicit, semi-implicit and implicit well-balanced schemes described in the paper are considered and the following acronyms are used:

- EXWBM1: explicit well-balanced numerical method of order one where the well-balanced reconstruction operator is based on RK collocation methods.
- SIWBM1: semi-implicit well-balanced numerical method of order one where the well-balanced reconstruction operator is based on RK collocation methods. Here, only the friction term is treated implicitly.
- IWBM1: implicit well-balanced numerical method of order one where the well-balanced reconstruction operator is based on RK collocation methods.

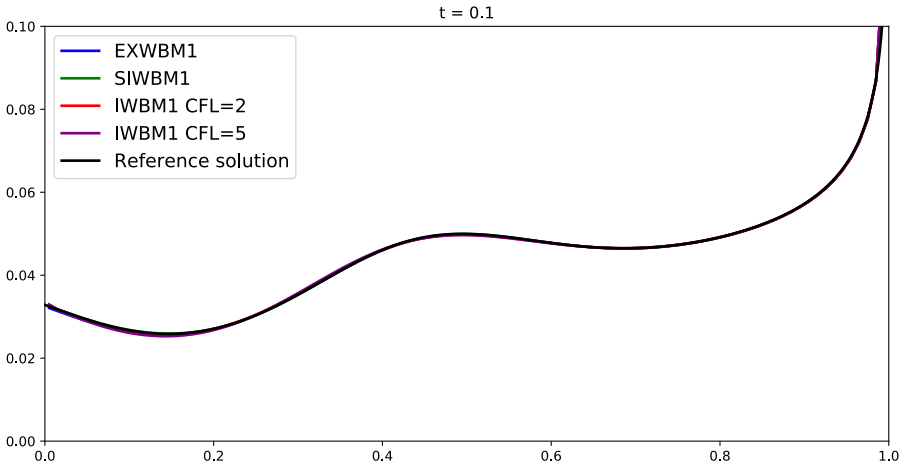
A reference solution has been computed using EXWBM1 with CFL=0.1 for a 1600-cell mesh. We have considered a mesh with 100 cells. Here, the explicit method requires a CFL value lower than 0.4 not to blow up, whereas semi-implicit and fully implicit methods admit larger values of CFL. Figure (2.5) shows the numerical solution for h and u at time the $t = 0.1$ for the following choices of the CFL value:

- CFL=0.3 has been considered for EXWBM1.
- CFL=0.9 has been considered for SIWBM1.
- CFL=2 and CFL=5 have been considered for IWBM1.

Notice that, with the previous choices, similar results have been obtained using the different schemes.



(a) h



(b) u

Figure 2.5: Shallow water with large value of the Manning friction coefficient: perturbation of a stationary solution. Reference and numerical solutions at $t = 0.1$ with a 100-cell mesh.

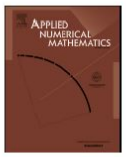
- **Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws**
  I. Gómez-Bueno, S. Boscarino, M.J. Castro, C. Parés and G. Russo. Applied Numerical Mathematics, 184 (2023): 18-48 (2023). DOI: `https://doi.org/10.1016/j.apnum.2022.09.016`.

Applied Numerical Mathematics 184 (2023) 18–48

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum



Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws

I. Gómez-Bueno [a,*], S. Boscarino [b], M.J. Castro [a], C. Parés [a], G. Russo [b]

[a] Departamento de Análisis Matemático, Estadística e I.O. y Matemática Aplicada, Facultad de Ciencias, Campus de Teatinos, Universidad de Málaga, 29071 Málaga, Spain
[b] Dipartimento di Matematica ed Informatica, Viale Andrea Doria, 6, 95125, University of Catania, Catania, Italy

ARTICLE INFO

Article history:
Received 12 July 2022
Received in revised form 11 September 2022
Accepted 24 September 2022
Available online 30 September 2022

Keywords:
Systems of balance laws
Well-balanced methods
Finite-volume methods
High-order methods
Reconstruction operators
Implicit methods
Semi-implicit methods
Shallow water equations

ABSTRACT

The aim of this work is to design implicit and semi-implicit high-order well-balanced finite-volume numerical methods for 1D systems of balance laws. The strategy introduced by two of the authors in some previous papers for explicit schemes based on the application of a well-balanced reconstruction operator is applied. The well-balanced property is preserved when quadrature formulas are used to approximate the averages and the integral of the source term in the cells. Concerning the time evolution, this technique is combined with a time discretization method of type RK-IMEX or RK-implicit. The methodology will be applied to several systems of balance laws.

Conclusions and future work

This thesis by compilation ends with a summary of the main contributions performed in the manuscripts which support it. Additionally, some possible future works that could arise from it are suggested.

Conclusions

The main goal of this thesis is the design of high-order well-balanced finite volume numerical methods for systems of balance laws of the form

$$U_f + f(U)_x = S(U)H_x,$$

where H is a known continuous function. We follow the proposal developed in [32], based on well-balanced reconstruction operators which are built from standard ones following a three-step general procedure. The main difficulty of this methodology arises in the first step, where local nonlinear problems consisting in finding a stationary solution in the stencil of the corresponding cell with prescribed average in the cell have to be solved. Then, the expression of the stationary solutions is needed, either in explicit or implicit form: the numerical methods so obtained are called exactly well-balanced.

The first three works which support this thesis are actually focused on this first step of the well-balanced reconstruction procedure, dealing with the design of explicit well-balanced methods for general systems of balance laws, regardless whether or not the expression of the stationary solutions is available. The first step of the reconstruction procedure has to be numerically solved, and the methods so obtained are not exactly well-balanced but just well-balanced. The local problems in the first step are redefined and the notion of discrete stationary solutions is introduced. The well-balanced property of these numerical methods is precisely stated. Two strategies are introduced to address these problems:

- A control-based approach. This strategy is based on the formulation of the local problems as control problems in functional form, where the control variable is in the initial condition of some Cauchy problems governed by the state equation. These control problems are solved by applying shooting-like methods, combined with

standard ODE solvers such as the fourth order RK method to solve the Cauchy problems for the state and adjoint systems. In particular, we consider two different strategies: Newton's method and descent methods with a suitable step (gradient and conjugate gradient methods). The performances of both techniques have been compared, concluding that the application of Newton's method is more efficient.

- The use of collocation RK methods. Since special care has to be taken if the control-based approach is chosen, due to the use of the fourth order RK method (which is not symmetric what may lead to the loss of the well-balanced property), this difficulty is circumvented by the application of the collocation RK methods to solve the local nonlinear problems in the first step.

Both strategies have been applied to a number of systems of balance laws, starting from simple scalar equations such as the Burgers equation with source term and gradually increasing the difficulty by considering more complex systems like the shallow water equations with topography and Manning friction or the compressible Euler equations with gravity. After performing several tests, we conclude that, although similar results are obtained with both strategies, the one based on the application of collocation RK methods is more efficient than the control-based approach. Moreover, it is important to highlight that it is the first time, to the best of our knowledge, that a family of high-order methods that preserve moving stationary solutions for Euler equations with gravity has been designed.

The proposals have been applied to tests cases where the stationary solutions are available, which allows us to compare the efficiency of the new implementation with respect to the exactly well-balanced methods in [32], with an affordable extra cost, but they have been also applied to equations where the only information about the steady states comes from the ODE

$$f(U)_x = S(U)H_x,$$

and thanks to which we can show the generality of the methods.

In addition, the results put on evidence that the well-balanced schemes are more effective than standard non well-balanced methods when they are applied to the propagation of small perturbation around an equilibrium or when long-time integration is required, despite being more costly. In any case, this extra computational cost is lower than the one required to obtain discretization errors of the order of machine precision by refining the mesh or increasing the order of non-well-balanced methods.

Moreover, a general technique that allows us to deal with resonant problems for any one-dimensional systems of balance laws is introduced, performing well for the case of the shallow water model. Two numerical tests for the Euler equations with gravity using first-order methods have been also included.

The last paper of this thesis is devoted to the development of implicit and semi-implicit high-order well-balanced methods. The key is to write the cell averages as the sum of

the cell averages at time t^n plus a time fluctuation. Moreover, the discrete reconstruction operator is the sum of the discrete well-balanced reconstruction operator at time t^n plus a standard reconstruction operator from the time fluctuations. Promising results for first- and second-order schemes have been obtained: several numerical tests are used to validate the schemes and to highlight the well-balanced property compliance.

Future works

- Implementation of third-order well-balanced schemes in the implicit and implicit-explicit cases.
- Comparison of strengths and weaknesses of the well-balanced methods introduced in this thesis and other such as the recent family of high-order and well-balanced methods for generic systems introduced in [76].
- Handling dry areas (for the shallow water equations) or vacuum (for the Euler equations). These are areas where a nonnegative variable vanishes, which causes important instabilities and may crash the code if the scheme is unable to handle them.
- Extension to the case where H has jump discontinuities, i.e., systems of balance laws with singular source terms. The solution U is expected to be discontinuous at a discontinuity of the function H, so that the source term $S(U)H_x$ is a nonconservative product that has to be defined. Our proposal is to start from the idea of Castro and Parés [32]: following the theory developed by Dal Maso, LeFloch and Murat in [83], the nonconservative products will be interpreted as Borel measures whose meaning depends on the choice of a family of paths, which must be consistent with the physics of the problem. The natural selection of the family of paths for systems of balance laws with singular source terms described in [32] will be applied.
- Development of asymptotic-preserving well-balanced (APWB) methods for hyperbolic systems that depend on a stiff relaxation parameter. A major difficulty arises when hyperbolic systems that depend on a parameter and converge to a limit system are approximated numerically. This limit system can also be of a different nature: it can happen, for example, that when this parameter tends to zero the behavior in the limit is parabolic, as occurs in the case of the hyperbolic heat equation or in the shallow water equations when the inverse of the friction coefficient tends to zero. The low Froude/low Mach number problems are also a challenge: for instance, the solutions of the equations for compressible flows converge to solutions of the equations for incompressible fluids as the Mach number tends to zero (see, for example, [127] and [128]). In such cases, it is of interest to have numerical methods that preserve the asymptotic behavior of the hyperbolic model (*asymptotic-preserving*): that is,

methods that, when the parameter tends to zero, are consistent with the asymptotic system and that are stable under the usual requirements for hyperbolic systems. When the wave propagation velocity tends to infinity as the parameter tends to zero, it is necessary to adopt implicit or semi-implicit methods so that they meet the stability requirement: see [84]. An additional difficulty in this case is to design numerical methods that are also well-balanced. Currently, *asymptotic-preserving* schemes are being developed for systems whose limit is a system of balance laws that are well-balanced for the limit problem when the relaxation parameter tends to zero. Prototypes that work well for academic problems are now available and are being verified on the models of interest.

- Extension to multidimensional systems. The difficulty increases considerably in this case because the stationary solutions no longer satisfy a system of ordinary differential equations, but in partial derivatives. Although RK methods cannot be easily extended to the solution of partial differential equations satisfied by the stationary solutions in the multidimensional case, in the particular case of collocation RK methods it is possible to extend their interpretation in terms of finding polynomials that satisfy the equation exactly at the quadrature points and that minimize a certain distance to the values of the approximations obtained in the stencil cells. The use of Discontinuous-Galerkin methods could be a good strategy, adopting a constrained optimization technique to look for approximated stationary solutions that are close to the stencil values and to ensure conservation in the cell. It is hoped that these strategies can lead to the development of schemes with well-balanced properties for general multidimensional equilibrium law systems. The development of methods of these characteristics is an open problem so any advance in this direction would have a high impact on the scientific community and a strong potential for applications.

Bibliography

[1] I. Gómez-Bueno, M. J. Castro, and C. Parés. High-order well-balanced methods for systems of balance laws: a control-based approach. *Applied Mathematics and Computation*, 394:125820, 2021.

[2] I. Gómez-Bueno, M.J. Castro, and C. Parés. Well-balanced reconstruction operator for systems of balance laws: Numerical implementation. In *Recent Advances in Numerical Methods for Hyperbolic PDE Systems*, pages 57–77. Springer, 2021.

[3] I. Gómez-Bueno, M.J. Castro, C. Parés, and G. Russo. Collocation methods for high-order well-balanced methods for systems of balance laws. *Mathematics*, 9(15):1799, 2021.

[4] I. Gómez-Bueno, S. Boscarino, M.J. Castro, C. Parés, and G. Russo. Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws. *Applied Numerical Mathematics*, 2022.

[5] R. Temam. *Navier-Stokes equations: theory and numerical analysis*, volume 343. American Mathematical Society, 2001.

[6] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[7] E. Gaburro, M. J. Castro, and M. Dumbser. Well-balanced Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming meshes for the euler equations of gas dynamics with gravity. *Monthly Notices of the Royal Astronomical Society*, 477(2):2251–2275, 2018.

[8] C. Klingenberg, G. Puppo, and M. Semplice. Arbitrary order finite volume well-balanced schemes for the Euler equations with gravity. *SIAM Journal on Scientific Computing*, 41(2):A695–A721, 2019.

[9] R. Manning, J.P. Griffith, T.F. Pigot, and L.F. Vernon-Harcourt. *On the flow of water in open channels and pipes*. 1890.

[10] P. Ripa. Conservation laws for primitive equations models with inhomogeneous layers. *Geophysical & Astrophysical Fluid Dynamics*, 70(1-4):85–111, 1993.

[11] P. Ripa. On improving a one-layer ocean model with thermodynamics. *Journal of Fluid Mechanics*, 303:169–201, 1995.

[12] S. Gottlieb and C.-W. Shu. Total variation diminishing Runge-Kutta schemes. *Mathematics of Computation*, 67(221):73–85, 1998.

[13] I. Cravero, G. Puppo, M. Semplice, and G. Visconti. CWENO: uniformly accurate reconstructions for balance laws. *Mathematics of Computation*, 87(312):1689–1719, 2018.

[14] C.-W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations*, pages 325–432. Springer, 1998.

[15] A. Harten, P. D. Lax, and B. Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM review*, 25(1):35–61, 1983.

[16] E. Haier, C. Lubich, and G. Wanner. *Geometric Numerical integration: structure-preserving algorithms for ordinary differential equations*. Springer, 2006.

[17] S. Boscarino and G. Russo. On a class of uniformly accurate imex runge–kutta schemes and applications to hyperbolic systems with relaxation. *SIAM Journal on Scientific Computing*, 31(3):1926–1945, 2009.

[18] S. Boscarino and G. Russo. Flux-explicit imex runge–kutta schemes for hyperbolic to parabolic relaxation problems. *SIAM Journal on Numerical Analysis*, 51(1):163–190, 2013.

[19] S. Boscarino, L. Pareschi, and G. Russo. Implicit-explicit runge–kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit. *SIAM Journal on Scientific Computing*, 35(1):A22–A51, 2013.

[20] I. Cordero-Carrión and P. Cerdá-Durán. Partially implicit runge-kutta methods for wave-like equations. In *Advances in Differential Equations and Applications*, pages 267–278. Springer, 2014.

[21] I. Cordero-Carrión. Partially implicit high order runge-kutta methods for wave-like equations in spherical-type coordinates. *Numerical Methods for Hyperbolic Equations*, page 211, 2012.

[22] V. Michel-Dansac and A. Thomann. Tvd-mood schemes based on implicit-explicit time integration. *Applied Mathematics and Computation*, 433:127397, 2022.

[23] V. Michel-Dansac and A. Thomann. On high-precision L^∞ -stable imex schemes for scalar hyperbolic multi-scale equations. In M.L. Muñoz-Ruiz, C. Parés, and G. Russo, editors, *Recent Advances in Numerical Methods for Hyperbolic PDE Systems*, pages 79–94, Cham, 2021. Springer International Publishing.

[24] G. Dimarco, R. Loubère, V. Michel-Dansac, and M.-H. Vignal. Second-order implicit-explicit total variation diminishing schemes for the euler system in the low mach regime. *Journal of Computational Physics*, 372:178–201, 2018.

[25] W. Boscheri, M. Tavelli, and C. E. Castro. An all froude high order imex scheme for the shallow water equations on unstructured voronoi meshes. *arXiv preprint arXiv:2209.00344*, 2022.

[26] W. Boscheri, M. Tavelli, and L. Pareschi. On the construction of conservative semi-lagrangian imex advection schemes for multiscale time dependent pdes. *Journal of Scientific Computing*, 90(3):1–46, 2022.

[27] W. Boscheri, M. Dumbser, M. Ioriatti, I. Peshkov, and E. Romenski. A structure-preserving staggered semi-implicit finite volume scheme for continuum mechanics. *Journal of Computational Physics*, 424:109866, 2021.

[28] V. Kučera, M. Lukáčová-Medviďová, S. Noelle, and J. Schütz. Low-mach consistency of a class of linearly implicit schemes for the compressible euler equations. *Programs and Algorithms of Numerical Mathematics*, pages 69–78, 2021.

[29] M. Lukáčová-Medvid'ová, J. Rosemeier, P. Spichtinger, and B. Wiebe. Imex finite volume methods for cloud simulation. In *International Conference on Finite Volumes for Complex Applications*, pages 179–187. Springer, 2017.

[30] G. Bispen, M. Lukáčová-Medvid'ová, and L. Yelash. Asymptotic preserving imex finite volume schemes for low mach number euler equations with gravitation. *Journal of Computational Physics*, 335:222–248, 2017.

[31] M. J. Castro, T. Morales de Luna, and C. Parés. Well-balanced schemes and path-conservative numerical methods. In *Handbook of Numerical Analysis*, volume 18, pages 131–175. Elsevier, 2017.

[32] M. J. Castro and C. Parés. Well-balanced high-order finite volume methods for systems of balance laws. *Journal of Scientific Computing, 82, 42*, 2020.

[33] J.M. Greenberg and AY. LeRoux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM Journal on Numerical Analysis*, 33(1):1–16, 1996.

[34] J.M. Greenberg, AY. Leroux, R. Baraille, and A. Noussair. Analysis and approximation of conservation laws with source terms. *SIAM Journal on Numerical Analysis*, 34(5):1980–2007, 1997.

[35] R. J. LeVeque. Balancing source terms and flux gradients in high-resolution godunov methods: The quasi-steady wave-propagation algorithm. *Journal of Computational Physics*, 146(1), 1998.

[36] L. Gosse. A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms. *Computers & Mathematics with Applications*, 39:135–159, 05 2000.

[37] B. Perthame and C. Simeoni. A kinetic scheme for the saint-venant system with a source term. *Calcolo*, 38:201–231, 11 2001.

[38] T. Chacón Rebollo, A. Dominguez Delgado, and E. D. Fernández-Nieto. A family of stable numerical solvers for the shallow water equations with source terms. *Computer Methods in Applied Mechanics and Engineering*, 192(1-2):203–225, 2003.

[39] H. Tang, T. Tang, and K. Xu. A gas-kinetic scheme for shallow-water equations with source terms. *Zeitschrift für angewandte Mathematik und Physik*, 55:365–382, 05 2004.

[40] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM Journal on Scientific Computing*, 25(6):2050–2065, 2004.

[41] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Springer Science & Business Media, 2004.

[42] T. Chacón Rebollo, A. Delgado, and E. Fernández-Nieto. Asymptotically balanced schemes for non-homogeneous hyperbolic systems - application to the shallow water equations. *Comptes Rendus Mathematique*, 338:85–90, 01 2004.

[43] 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. *Journal of Computational Physics*, 214:567–598, 05 2006.

[44] S. Noelle, N. Pankratz, G. Puppo, and J. R. Natvig. Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows. *Journal of Computational Physics*, 213(2):474–499, 2006.

[45] S. Noelle, Y. Xing, and C.-W. Shu. High-order well-balanced finite volume weno schemes for shallow water equation with moving water. *Journal of Computational Physics*, 226:29–58, 09 2007.

[46] M. J. Castro, T. Chacón Rebollo, E. D. Fernández-Nieto, and C. Parés. On well-balanced finite volume methods for nonconservative nonhomogeneous hyperbolic systems. *SIAM Journal on Scientific Computing*, 29(3):1093–1126, 2007.

[47] M. Lukáčová-Medvid'ová, S. Noelle, and M. Kraft. Well-balanced finite volume evolution galerkin methods for the shallow water equations. *Journal of Computational Physics*, 221:122–147, 01 2007.

[48] M. Pelanti, F. Bouchut, and A. Mangeney. A Roe-type scheme for two-phase shallow granular flows over variable topography. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(5):851–885, 2008.

[49] M. Dudzinski and M. Lukáčová-Medvid'ová. Well-balanced bicharacteristic-based scheme for multilayer shallow water flows including wet/dry fronts. *Journal of Computational Physics*, 235:82–113, 2013.

[50] A. Chertock, S. Cui, A. Kurganov, and T. Wu. Well-balanced positivity preserving central-upwind scheme for the shallow water system with friction terms. *International Journal for Numerical Methods in Fluids*, 78, 04 2015.

[51] Y.N. Cheng and A. Kurganov. Moving-water equilibria preserving central-upwind schemes for the shallow water equations. *Communications in Mathematical Sciences*, 14:1643–1663, 01 2016.

[52] V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher. A well-balanced scheme for the shallow-water equations with topography. *Computers & Mathematics with Applications*, 72(3):568–593, 2016.

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

[54] A. Kurganov. Finite-volume schemes for shallow-water equations. *Acta Numerica*, 27:289–351, 05 2018.

[55] A. Chertock, M. Dudzinski, A. Kurganov, and M. Lukáčová-Medvid'ová. Well-balanced schemes for the shallow water equations with coriolis forces. *Numerische Mathematik*, 138(4):939–973, 2018.

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

[57] L. Arpaia and M. Ricchiuto. Well balanced residual distribution for the ale spherical shallow water equations on moving adaptive meshes. *Journal of Computational Physics*, 405:109173, 2020.

[58] V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher. A two-dimensional high-order well-balanced scheme for the shallow water equations with topography and manning friction. *Computers & Fluids*, 230:105152, 2021.

[59] A. Kurganov, Y. Liu, and M. Lukáčová-Medvid'ová. A well-balanced asymptotic preserving scheme for the two-dimensional rotating shallow water equations with nonflat bottom topography. *SIAM Journal on Scientific Computing*, 44(3):A1655–A1680, 2022.

[60] M. Ciallella, D. Torlo, and M. Ricchiuto. Arbitrary high order weno finite volume scheme with flux globalization for moving equilibria preservation. *arXiv preprint arXiv:2205.13315*, 2022.

[61] Arpaia. L., M. Ricchiuto, Filippini A.G., and R. Pedreros. An efficient covariant frame for the spherical shallow water equations: Well balanced dg approximation and application to tsunami and storm surge. *Ocean Modelling*, 169:101915, 2022.

[62] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. Well-balanced schemes to capture non-explicit steady states: Ripa model. *Mathematics of Computation*, 85(300):1571–1602, 2016.

[63] L. Gosse. A well-balanced scheme using non-conservative products designed for hyperbolic systems of conservation laws with source terms. *Mathematical Models and Methods in Applied Sciences*, 11, 06 1999.

[64] R. Käppeli and S. Mishra. Well-balanced schemes for the euler equations with gravitation. *Journal of Computational Physics*, 259:199 – 219, 2014.

[65] P. Chandrashekar and M. Zenk. Well-balanced nodal discontinuous galerkin method for euler equations with gravity. *Journal of Scientific Computing*, 71, 11 2015.

[66] C. Klingenberg, R. Touma, and U. Koley. Well-balanced unstaggered central schemes for the euler equations with gravitation. *SIAM Journal on Scientific Computing*, 38, 01 2016.

[67] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. A well-balanced scheme to capture non-explicit steady states in the Euler equations with gravity. *International Journal for Numerical Methods in Fluids*, 81(2):104–127, 2016.

[68] R. Käppeli and S. Mishra. A well-balanced finite volume scheme for the euler equations with gravitation-the exact preservation of hydrostatic equilibrium with arbitrary entropy stratification. *Astronomy & Astrophysics*, 587:A94, 2016.

[69] D. Varma and P. Chandrashekar. A second-order well-balanced finite volume scheme for euler equations with gravity. *Computers & Fluids*, 181, 03 2018.

[70] E. Franck and L. S. Mendoza. Finite volume scheme with local high order discretization of the hydrostatic equilibrium for the euler equations with external forces. *Journal of Scientific Computing*, 69(1):314–354, 2016.

[71] E. Gaburro, M. J. Castro, and M. Dumbser. Well-balanced arbitrary-lagrangian-eulerian finite volume schemes on moving nonconforming meshes for the euler equations of gas dynamics with gravity. *Monthly Notices of the Royal Astronomical Society*, 477(2):2251–2275, 2018.

[72] X. Liu, A. Chertock, and A. Kurganov. An asymptotic preserving scheme for the two-dimensional shallow water equations with coriolis forces. *Journal of Computational Physics*, 391:259–279, 2019.

[73] F. Kanbar, R. Touma, and C. Klingenberg. Well-balanced central schemes for the one and two-dimensional euler systems with gravity. *Applied Numerical Mathematics*, 156, 06 2020.

[74] L. Grosheintz-Laval and R. Käppeli. Well-balanced finite volume schemes for nearly steady adiabatic flows. *Journal of Computational Physics*, 423:109805, 2020.

[75] M. V. Popov, R. Walder, D. Folini, T. Goffrey, I. Baraffe, T. Constantino, C. Geroux, J. Pratt, M. Viallet, and R. Käppeli. A well-balanced scheme for the simulation tool-kit a-maze: implementation, tests, and first applications to stellar structure. *Astronomy & Astrophysics*, 630:A129, 2019.

[76] C. Berthon, S. Bulteau, F. Foucher, M. M'Baye, and V. Michel-Dansac. A very easy high-order well-balanced reconstruction for hyperbolic systems with source terms. *SIAM Journal on Scientific Computing*, 44(4):A2506–A2535, 2022.

[77] R. Abgrall and M. Ricchiuto. Hyperbolic balance laws: residual distribution, local and global fluxes. *Numerical Fluid Dynamics*, pages 177–222, 2022.

[78] A. Bermúdez and M. E. Vázquez-Cendón. Upwind methods for hyperbolic conservation laws with source terms. *Computers & Fluids*, 23(8):1049–1071, 1994.

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

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

[81] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2013.

[82] E. Cerdá Tena. Optimización dinámica. *Prentecie may*, 2001.

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

[84] S. Boscarino, P. G. LeFloch, and G. Russo. High-order asymptotic-preserving methods for fully nonlinear relaxation problems. *SIAM Journal on Scientific Computing*, 36(2):A377–A395, 2014.

[85] E. Godlewski and P. A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Springer, 1995.

[86] R. J. LeVeque. *Numerical methods for conservation laws*, volume 132. Springer, 1992.

[87] T. P. Liu. The Riemann problem for general systems of conservation laws. *Journal of Differential Equations*, 18(1):218–234, 1975.

[88] T. P. Liu. The deterministic version of the Glimm scheme. *Communications in Mathematical Physics*, 57(2):135–148, 1977.

[89] J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Communications on Pure and Applied Mathematics*, 18(4):697–715, 1965.

[90] P. D. Lax. Hyperbolic systems of conservation laws II. *Communications on Pure and Applied Mathematics*, 10(4):537–566, 1957.

[91] P. D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, volume 11. Society for Industrial and Applied Mathematics, 1973.

[92] P. G. LeFloch. *Hyperbolic Systems of Conservation Laws: The theory of classical and nonclassical shock waves*. Springer Science & Business Media, 2002.

[93] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Springer Science & Business Media, 2013.

[94] M. Dumbser and M. Käser. Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems. *Journal of Computational Physics*, 221(2):693–723, 2007.

[95] 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. *Journal of Computational Physics*, 226(1):204–243, 2007.

[96] 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. *Journal of Computational Physics*, 227(18):8209–8253, 2008.

[97] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, III. In *Upwind and high-resolution schemes*, pages 218–290. Springer, 1987.

[98] A. Marquina. Local piecewise hyperbolic reconstruction of numerical fluxes for nonlinear scalar conservation laws. *SIAM Journal on Scientific Computing*, 15(4):892–915, 1994.

[99] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. In *Upwind and High-Resolution Schemes*, pages 328–374. Springer, 1989.

[100] D. Levy, G. Puppo, and G. Russo. Central WENO schemes for hyperbolic systems of conservation laws. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 33(3):547–571, 1999.

[101] D. Levy, G. Puppo, and G. Russo. Compact central WENO schemes for multidimensional conservation laws. *SIAM Journal on Scientific Computing*, 22(2):656–672, 2000.

[102] B. Van Leer. Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme. *Journal of Computational Physics*, 14(4):361–370, 1974.

[103] B. Van Leer. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *Journal of Computational Physics*, 23(3):276–299, 1977.

[104] B. Van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *Journal of Computational Physics*, 32(1):101–136, 1979.

[105] S. Serna and A. Marquina. Power eno methods: a fifth-order accurate weighted power eno method. *Journal of Computational Physics*, 194(2):632–658, 2004.

[106] G. S. Jiang and C.-W. Shu. Efficient implementation of weighted ENO schemes. *Journal of Computational Physics*, 126(1):202–228, 1996.

[107] A. Baeza, R. Bürger, P. Mulet, and D. Zorío. On the efficient computation of smoothness indicators for a class of WENO reconstructions. *Journal of Scientific Computing*, 80(2):1240–1263, 2019.

[108] H. Carrillo, C. Parés, and D. Zorío. Lax-Wendroff approximate taylor methods with fast and optimized weighted essentially non-oscillatory reconstructions. *Journal of Scientific Computing*, 86(1):1–41, 2021.

[109] I. Cravero and M. Semplice. On the accuracy of WENO and CWENO reconstructions of third order on nonuniform meshes. *Journal of Scientific Computing*, 67(3):1219–1246, 2016.

[110] G. Puppo and M. Semplice. Well-balanced high order 1D schemes on non-uniform grids and entropy residuals. *Journal of Scientific Computing*, 66(3):1052–1076, 2016.

[111] V. A. Titarev and E. F. Toro. Ader: Arbitrary high order godunov approach. *Journal of Scientific Computing*, 17(1):609–618, 2002.

[112] E. F. Toro and V. A. Titarev. Solution of the generalized Riemann problem for advection-reaction equations. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 458(2018):271–281, 2002.

[113] V. A. Titarev and E. F. Toro. ADER schemes for three-dimensional non-linear hyperbolic systems. *Journal of Computational Physics*, 204(2):715–736, 2005.

[114] E. F. Toro and V. A. Titarev. Derivative Riemann solvers for systems of conservation laws and ADER methods. *Journal of Computational Physics*, 212(1):150–165, 2006.

[115] M. Dumbser and C.-D. Munz. Building blocks for arbitrary high order discontinuous Galerkin schemes. *Journal of Scientific Computing*, 27(1-3):215–230, 2006.

[116] M. Dumbser, C. Enaux, and E. F. Toro. Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *Journal of Computational Physics*, 227(8):3971–4001, 2008.

[117] G. Li, L. Song, and J. Gao. High order well-balanced discontinuous galerkin methods based on hydrostatic reconstruction for shallow water equations. *Journal of Computational and Applied Mathematics*, 340:546–560, 2018.

[118] 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. *Applied Mathematics and Computation*, 395:125848, 2021.

[119] 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. *Computers & Mathematics with Applications*, 82:179–199, 2021.

[120] M. Dumbser, M. J. Castro, C. Parés, and E. F. Toro. ADER schemes on unstructured meshes for nonconservative hyperbolic systems: Applications to geophysical flows. *Computers & Fluids*, 38(9):1731–1748, 2009.

[121] N. Izem, M. Seaid, and M. Wakrim. A discontinuous galerkin method for two-layer shallow water equations. *Mathematics and Computers in Simulation*, 120:12–23, 2016.

[122] C. Escalante, T. Morales de Luna, and M. J. Castro. Non-hydrostatic pressure shallow flows: Gpu implementation using finite volume and finite difference scheme. *Applied Mathematics and Computation*, 338:631–659, 2018.

[123] D. Zorío, A. Baeza, and P. Mulet. An approximate Lax–Wendroff-type procedure for high order accurate schemes for hyperbolic conservation laws. *Journal of Scientific Computing*, 71(1):246–273, 2017.

[124] H. Carrillo and C. Parés. Compact approximate taylor methods for systems of conservation laws. *Journal of Scientific Computing*, 80(3):1832–1866, 2019.

[125] M. J. Castro, J. M. Gallardo, J. A. López-García, and C. Parés. Well-balanced high order extensions of Godunov's method for semilinear balance laws. *SIAM Journal on Numerical Analysis*, 46(2):1012–1039, 2008.

[126] P. Wolfe. Convergence conditions for ascent methods. *SIAM review*, 11(2):226–235, 1969.

[127] H. Guillard and C. Viozat. On the behaviour of upwind schemes in the low mach number limit. *Computers & fluids*, 28(1):63–86, 1999.

[128] C. Chalons, M. Girardin, and S. Kokh. An all-regime lagrange-projection like scheme for 2d homogeneous models for two-phase flows on unstructured meshes. *Journal of Computational Physics*, 335:885–904, 2017.