

SOLUCIÓN NUMÉRICA DE LAS
ECUACIONES DE AGUAS SOMERAS EN
CANALES RECTANGULARES CON
GALERKIN DISCONTINUO

TESIS

QUE PARA OBTENER EL TÍTULO DE:
**Maestro en Ciencias (Matemáticas
Aplicadas e Industriales)**

PRESENTA:

José Carlos Sánchez Fernández

Asesora:

Dra. María Luisa Sandoval Solís

Jurado calificador:

Presidente: Dr.

Secretario: Dr.

Vocal: Dr.

Ciudad de México, 2017

Índice general

1. Introducción	1
1.1. Objetivos de la tesis	4
1.2. Objetivos específicos	4
1.3. Estructura de la tesis	4
2. Las ecuaciones de aguas someras	7
2.1. Derivación de las ecuaciones de aguas someras	7
2.2. Leyes de conservación	8
2.3. Ecuaciones sobre una superficie libre	8
2.3.1. Condiciones de frontera	9
2.4. Ecuaciones de aguas someras	10
2.5. Propiedades de las ecuaciones de aguas someras	13
2.6. El problema de Riemann para las ecuaciones de aguas someras	17
3. Método de Galerkin discontinuo para el problema de estudio	27
3.1. Ecuaciones de Saint-Venant	27
3.2. Discretización del dominio	28
3.3. Definición de espacios y normas	30
3.4. Formulación variacional en espacios de Sobolev particionados	31
3.5. Discretización del problema	31
3.6. Estimación del error para el método de Galerkin discontinuo	35
3.7. Funciones de forma	35
3.8. Transformaciones isoparamétricas	36
3.9. Integración en el espacio	37
3.10. Flujos numéricos	41

3.10.1. Esquema de Godunov	41
3.10.2. <i>Solvers</i> de Riemann	44
4. Integración en el tiempo	53
4.1. Esquemas numéricos para integrar el tiempo	53
4.1.1. Método de Euler	53
4.1.2. Método de Runge-Kutta de segundo orden	54
4.2. Esquemas TVD	55
4.3. Limitadores de pendiente	58
5. Resultados Numéricos	61
5.1. Análisis de limitadores y flujos numéricos	61
5.1.1. Ecuación de convección	61
5.1.2. Ecuación de Burgers	69
5.2. Ecuaciones de aguas someras	76
5.2.1. Fractura de presa idealizada en un canal rectangular sobre lecho mojado	76
5.2.2. Fractura de presa idealizada en un canal rectangular sobre lecho seco	77
5.2.3. Fractura de presa en un canal artificial rectangular con fricción sobre lecho seco	79
5.2.4. Flujo sobre bol parabólico	80
5.2.5. Salto hidráulico	82
5.2.6. Flujo sobre un tope	83
5.2.7. Flujo sobre un lecho irregular	85
6. Conclusiones y trabajo futuro	87
6.1. Conclusiones	87
6.2. Trabajo futuro	88
A. Ecuaciones hiperbólicas	89
A.1. Ecuaciones hiperbólicas escalares de primer orden	89
A.1.1. Ecuaciones lineales y el método de las características	89
A.1.2. Ecuaciones cuasilineales	92
A.2. Formulación débil de ecuaciones escalares de primer orden	96
A.3. El problema de Riemann	97

A.4. Sistemas de primer orden	100
A.4.1. Sistemas lineales	100
A.4.2. Sistemas cuasilineales	102
A.4.3. Método de las características	104
A.5. Formulación débil de sistemas de primer orden	105
A.5.1. El problema de Riemann	106
B. Condiciones de frontera	111
B.1. Condiciones de frontera: Caso lineal	111
B.2. Condiciones de frontera: Caso no lineal	113
B.3. Condiciones de frontera: Ecuaciones de aguas someras	115
Bibliografía	118

Capítulo 1

Introducción

El flujo de agua en canales es un fenómeno que tiene una variada aplicación en problemas importantes tales como inundaciones, flujos de ríos, canales de riego, vaciado de depósitos, rotura de presas, etcétera. Específicamente, el tema de inundaciones es una de las motivaciones de esta tesis. Con esto en mente se necesita primero entender el modelo que describa el movimiento del agua en un cauce y generar una metodología robusta para simularlo numéricamente, lo cual es el propósito de este trabajo.

En este trabajo utilizamos las ecuaciones de Saint-Venant o de aguas someras unidimensionales para modelar los cambios de caudal y profundidad de un líquido a lo largo de un canal abierto o de superficie libre. Estas relaciones son un conjunto de ecuaciones diferenciales hiperbólicas y pueden interpretarse como una contracción de las ecuaciones de aguas someras en 2 dimensiones.

Las ecuaciones de aguas poco profundas fueron desarrolladas por el matemático e ingeniero francés Adhémar Jean Claude Barré de Saint-Venant en 1871 y vienen de las ecuaciones de conservación de masa y de momento lineal. Bajo las suposiciones de que el flujo es incompresible y que la presión es hidrostática, se pueden derivar de la integración vertical de las ecuaciones de Navier-Stokes.

Cabe mencionar que son extensamente ocupadas en modelos computacionales dado que son significativamente más fáciles de resolver que la versión en dos o tres dimensiones. Su aplicación va desde flujo de ríos hasta análisis de rotura de presa o salto hidráulico. Por estas características es un modelo adecuado para generar un software robusto en 1D. Sin embargo en un futuro se emplearán modelos 2D y 3D adaptando la metodología desarrollada.

Cuando se desea resolver numéricamente una ecuación diferencial parcial o un sistema de ellas se puede escoger algún método de la gran variedad que existen, entre los que se encuentran diferencias finitas, elementos finitos y volúmenes finitos. Todos estos suelen usarse en la solución de la parte espacial. Si además se desea resolver un problema no estacionario existen otra diversidad de técnicas empleadas para tratar esto. Entonces, la pregunta se reduce a que combinación de estas técnicas es la adecuada. Como es de esperarse, cada uno tiene ventajas y desventajas. Ver (Hesthaven y Warburton, 2007; Toro, 2001)

Por ejemplo, el método de volumen finito (VF) aproxima la solución sobre un conjunto de elementos que discretizan el espacio suponiendo que en el centro de cada elemento la solución es constante. Esto permite que el método sea capaz de trabajar con mallas complejas. Así mismo, es necesario el cálculo de los flujos numéricos en las interfaces de los elementos, llevando a distintos tipos de métodos de volúmenes finitos. Sin embargo este método no permite que el orden de precisión aumente de manera sencilla; de hacerlo se pierde la flexibilidad geométrica de la malla.

Otro tipo de esquema es el método de elemento finito (EF), el cual usa la formulación variacional del problema original y toma la idea de discretizar el espacio en elementos, con la diferencia de que en cada uno de ellos la solución es aproximada por una base de funciones. Esto permite un alto orden de precisión simplemente aumentando el número de las funciones base. No obstante, necesita construir un sistema algebraico de ecuaciones que puede ser muy costoso de resolver.

Por su parte, el método de Galerkin discontinuo (GD) es una clase de método de elemento finito. Al igual que en EF, GD busca una aproximación a la solución de la formulación variacional del problema original. Esta aproximación se obtiene tomando la proyección de la solución buscada sobre un espacio de dimensión finita, cuya base permite funciones discontinuas que generalmente son polinomios a trozos. Por otro lado, la manera de tratar las interfaces de los elementos está ligada a los métodos de volúmenes finitos, ya que se emplean los flujos numéricos o *solvers* de Riemann desarrollados para VF. De esta manera, GD puede verse como una combinación de los métodos de EF y VF.

Debido a lo anterior, Galerkin Discontinuo tiene cierta ventaja sobre el método de EF: la posibilidad de mallados arbitrarios (mallas no conformes), completa libertad de cambiar los grados de los polinomios en cada elemento independientemente de aquellos en los vecinos, estructura de datos extremadamente local (está enfocada en el elemento y sus fronteras), no es necesario ensamblar una matriz global, liberando la demanda de memoria de disco y facilidad de paralelizar el método.

Sin embargo, como es de esperarse, también tiene desventajas. Una de ellas es el aumento de grados de libertad, causada por mantener el esquema local. Por ejemplo, en el caso de elementos lineales, los grados de libertad se duplican. Además, debido a que las funciones base son discontinuas, los términos que involucran segundas derivadas se deben tratar con métodos especiales que necesariamente llevan a un aumento del número de incógnitas. Una discusión más amplia sobre las ventajas y desventajas de este método puede consultarse en (Li, 2005)

Por último, señalamos que la técnica de Galerkin discontinuo se ha escogido para el desarrollo de esta tesis, debido a que puede capturar ondas de choque y de rarefacción que aparecen frecuentemente en la solución de las ecuaciones de Saint-Venant, así como posibles soluciones con discontinuidades o gradientes pronunciados (declives muy marcados). Para lograr esto y generar un software robusto se prueban los métodos de Euler hacia adelante y de Runge-Kutta de segundo orden para la integración en el tiempo. También se reescriben dichos esquemas para que satisfagan la propiedad de disminuir la variación total (TVD) y

se compara su desempeño con las versiones originales, al igual que en VF se requiere incluir limitadores de pendiente para disminuir las oscilaciones espurias que aparecen en zonas no diferenciables de la solución. Este trabajo sigue lo expuesto en ((Cockburn et al., 1989; Cockburn y Shu, 1989; Hesthaven y Warburton, 2007; Khan y Lai, 2014).

Breve historia del método

El método de Galerkin discontinuo se presentó por primera vez en (Reed y Hill, 1973) para resolver la ecuación estacionaria de transporte de neutrones. La solución obtenida se calculó elemento por elemento y éstos se ordenaron en la dirección de las características debida a la naturaleza lineal del problema.

En (LeSaint y Raviart, 1974) se realizó el primer análisis numérico del método, donde demuestran convergencia $O(h^N)$ sobre una malla general con triángulos y convergencia $O(h^{N+1})$ para una malla con cuadriláteros de tamaño h usando una aproximación local por polinomios de grado N . Más tarde en (Johnson y Pitkäranta, 1986) se probó que el método tiene una tasa de convergencia de $h^{N+1/2}$ para mallas generales. En (Godlewski y Raviart, 1996) y en (Hesthaven y Warburton, 2007) se establecen cotas para el error en los espacios de Sobolev particionados.

Las primeras aplicaciones del método involucraron la propagación de ondas en medios elásticos (Wellford y Oden, 1975), modelado de ecuaciones parabólicas (Jamet, 1978) y simulación de flujos viscoelásticos (Fortin y Fortin, 1989). En (Chavent y Salzano, 1982) se aplicó para resolver leyes de conservación no lineales unidimensionales, con elementos lineales en el espacio y el método de Euler hacia adelante para la integración temporal. El esquema resultó ser inestable a pesar de usar un paso de tiempo muy restrictivo. Para remediar esto, en (Chavent y Cockburn, 1989) se introdujo un limitador de pendiente adecuado, obteniendo un esquema que disminuya la variación total en las medias (TVDM por sus siglas en inglés) y un esquema con variación total acotada (TVB). Estos esquemas fueron estables para números de Courant-Friedrichs-Lewy menores a 1/2. Pero estos esquemas solo fueron de primer orden y el limitador afectó la calidad de la solución en regiones suaves. Este problema fue corregido en (Cockburn y Shu, 1991) al utilizar el método de Runge-Kutta de segundo orden TVD para integrar el tiempo y un limitador de pendiente mejorado. Este esquema resultó ser linealmente estable para números CFL menores a 1/3, manteniendo la exactitud en regiones suaves, asegurando soluciones discontinuas sin oscilaciones y convergencia a la solución de entropía. Más tarde, en (Cockburn y Shu, 1989) se extendió este resultado a métodos de Runge-Kutta de mayor orden para leyes de conservación con limitadores de pendiente generales. Además estos resultados se ampliaron a sistemas unidimensionales, (Cockburn et al., 1989) y multidimensionales (Cockburn et al., 1990).

1.1. Objetivos de la tesis

El objetivo general de la tesis es simular el flujo de agua en canales rectangulares con cama irregular y fricción y sección transversal homogénea.

1.2. Objetivos específicos

Con el fin de entender el problema y la metodología se consideran los siguientes objetivos específicos:

1. Estudio de la teoría de las ecuaciones hiperbólicas, problema de Riemann, condiciones de Rankine-Hugoniot y condiciones de entropía, sistemas lineales y cuasilineales de ecuaciones hiperbólicas y condiciones de frontera.
2. Derivación de las ecuaciones de Saint-Venant en 1D y 2D.
3. Descripción del método de Galerkin Discontinuo, de los esquemas TVD y de los limitadores de pendiente.
4. Tratamiento del problema seco-mojado.
5. Realizar diferentes pruebas:
 - a) Rotura de presa sobre cama mojada o seca, con y sin fricción.
 - b) Salto hidráulico.
 - c) Flujo de agua sobre un bol parabólico.
 - d) **Diversos regímenes de flujo de agua sobre un tope.**
 - e) **Flujo de agua sobre una cama irregular.**

1.3. Estructura de la tesis

El presente documento se organiza en 6 capítulos. En el primero se presenta una introducción al tema de estudio, las ideas principales y una breve historia del método de Galerkin discontinuo y de los métodos TVD que se utilizaron. Conjuntamente aparecen los objetivos de la tesis. En el segundo capítulo se derivan las ecuaciones de aguas someras y la solución analítica de ellas para el caso homogéneo.

El capítulo tres muestra la discretización de los modelos por medio del esquema de Galerkin discontinuo, incluyendo las funciones de forma, las transformadas isoparámetricas y los flujos numéricos. Los métodos de integración en el tiempo se describen en el capítulo cuatro

junto con las propiedades TVD, TVB y TVBM de los métodos de integración temporal, así como la formula del limitador de pendiente.

El capítulo cinco engloba las simulaciones desarrolladas para las ecuaciones de aguas poco profundas en varios escenarios. El último capítulo incluye las conclusiones del trabajo y las posibles ideas para un trabajo futuro.

En los anexos, se presenta el método de las características y las propiedades que deben presentar las condiciones de frontera para ecuaciones hiperbólicas.

Capítulo 2

Las ecuaciones de aguas someras

En este capítulo se detalla el modelo físico a resolver, derivándolo a partir de las ecuaciones de conservación de la masa y el momento tridimensional. Se muestra que con el fin de reducir el número de variables, tanto independientes como dependientes, se utilizan algunas simplificaciones para llegar al problema de flujo en una superficie libre en dos dimensiones. Finalmente, se presentan las ecuaciones con las que se trabajará en todo este documento, es decir, las ecuaciones de aguas someras unidimensionales.

Por otro lado, con motivo de comparar las soluciones obtenidas numéricamente en el problema de rotura de presa, se desarrolla la solución analítica para el problema de Riemann con estas ecuaciones en el caso homogéneo. Para lograrlo se aplica el método de las características y las invariantes de Riemann. Más detalles sobre este método se puede consultar en el apéndice A y en (Mattheij et al., 2005).

2.1. Derivación de las ecuaciones de aguas someras

La derivación de las ecuaciones de aguas someras está basada en lo expuesto en (Toro, 2001). Este análisis necesita de las suposiciones siguientes:

- a) No hay erosión ni acreción en el lecho.
- b) La densidad es constante y el fluido incompresible.
- c) Sólo actúa la fuerza de gravedad sobre el agua, es decir, la presión es hidrostática.
- d) La pendiente promedio en el lecho del canal se considera pequeña.
- e) La presión atmosférica se considera igual a cero por simplicidad.
- f) La componente vertical de la aceleración es despreciable.
- g) Las variables dependientes se consideran continuas y diferenciables.

2.2. Leyes de conservación

Primeramente se presentan las ecuaciones para las leyes de conservación de la masa y el momento en materiales compresibles; éstas se escriben como sigue:

$$\rho_t + \nabla \cdot (\rho \mathbf{V}) = 0, \quad (2.1a)$$

$$\frac{\partial}{\partial t} (\rho \mathbf{V}) + \nabla \cdot [\rho \mathbf{V} \otimes \mathbf{V} + p \mathbf{I} - \Pi] = \rho \mathbf{g}.^1 \quad (2.1b)$$

Las variables independientes son t el tiempo y x, y, z en el espacio. Las dependientes son ρ la densidad, p la presión y $\mathbf{V} = (u, v, w)^T$ la velocidad, con u, v, w las componentes de la velocidad en las direcciones x, y, z , respectivamente. El vector $\mathbf{g} = (g_1, g_2, g_3)^T$ representa las fuerzas que actúan sobre el material. Además, Π es la matriz de tensión viscosa, que se supondrá despreciable. Por último, \mathbf{I} es la matriz identidad. A partir de estas ecuaciones se deducirán las ecuaciones de aguas someras utilizadas en este trabajo.

2.3. Ecuaciones sobre una superficie libre

Al escribir las ecuaciones (2.1) en términos de las variables físicas, considerando que la densidad es constante, obtenemos

$$u_x + v_y + w_z = 0, \quad (2.2a)$$

$$u_t + uu_x + vv_y + ww_z + \frac{1}{\rho} p_x = g_1, \quad (2.2b)$$

$$v_t + uv_x + vv_y + ww_z + \frac{1}{\rho} p_y = g_2, \quad (2.2c)$$

$$w_t + uw_x + vw_y + ww_z + \frac{1}{\rho} p_z = g_3. \quad (2.2d)$$

Se considera el flujo de agua con una superficie libre, bajo la gravedad en un dominio tridimensional. Las variables espaciales x, y denotan el plano horizontal y z la dirección vertical. El fondo del canal se define por medio de la función

$$z = b(x, y) \quad (2.3)$$

y la superficie es descrita por

$$z = s(x, y, t) = b(x, y) + h(x, y, t), \quad (2.4)$$

donde $h(x, y, t)$ es la profundidad del agua en el punto (x, y) y al tiempo t , esto es, la

¹La divergencia de este tensor se define como sigue: $\nabla \cdot [\rho \mathbf{V} \otimes \mathbf{V}] = (\nabla \cdot \rho \mathbf{V}) \mathbf{V} + (\rho \mathbf{V} \cdot \nabla) \mathbf{V}$.

distancia entre el fondo y la superficie libre. La figura 2.1 muestra estas variables en un corte longitudinal al canal para cierto valor de y .

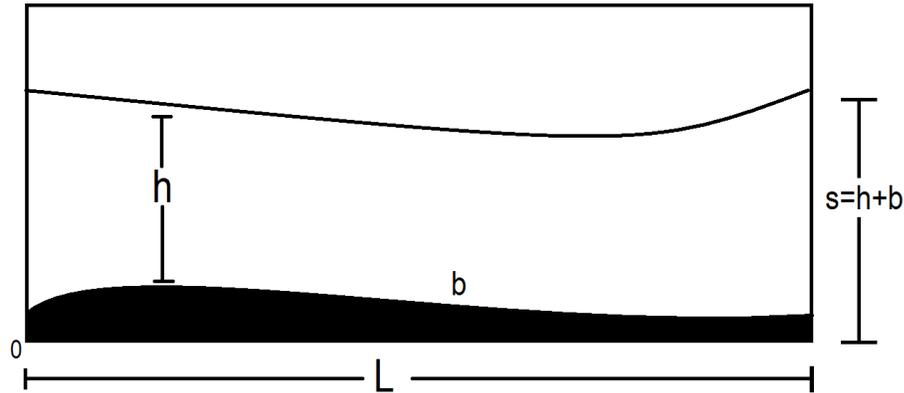


Figura 2.1: Flujo sobre una superficie libre.

Suponiendo que el vector de fuerzas que actúan sobre el agua es $\mathbf{g} = (0, 0, -g)^T$ con $g = 9.8 \text{ m/s}^2$ la aceleración de la gravedad, las ecuaciones (2.2) se convierten en

$$u_x + v_y + w_z = 0, \quad (2.5a)$$

$$u_t + uu_x + vv_y + ww_z = -\frac{1}{\rho}p_x, \quad (2.5b)$$

$$v_t + uv_x + vv_y + vw_z = -\frac{1}{\rho}p_y, \quad (2.5c)$$

$$w_t + uw_x + vw_y + ww_z = -\frac{1}{\rho}p_z - g. \quad (2.5d)$$

Este sistema de ecuaciones tendrá solución cuando se impongan condiciones iniciales y de frontera adecuadas. Sin embargo, su solución seguirá teniendo un alto grado de dificultad. En específico porque la frontera correspondiente a la superficie varía con el tiempo.

2.3.1. Condiciones de frontera

Es necesario especificar que forma tienen las condiciones de frontera para el sistema de ecuaciones (2.5). Para este propósito, sea $\Psi(x, y, z, t) = 0$ alguna de las fronteras. Así, para la superficie libre del agua

$$\Psi(x, y, z, t) = z - s(x, y, t) = 0, \quad (2.6)$$

y para el fondo

$$\Psi(x, y, z, t) = z - b(x, y) = 0. \quad (2.7)$$

A la relación (2.6) le imponemos dos condiciones, La primera de éstas es la condición cinemática

$$\frac{d\Psi}{dt} = \Psi_t + u\Psi_x + v\Psi_y + w\Psi_z = 0,$$

que indica que las partículas del fluido en la superficie libre siempre permanecen sobre la superficie libre. Así

$$(-s_t - us_x - vs_y + w)|_{z=s} = 0. \quad (2.8)$$

La segunda es la condición dinámica

$$p(x, y, z, t)|_{z=s} - p_{atm} = 0, \quad (2.9)$$

donde p_{atm} es la presión atmosférica, e indica que la presión en la superficie libre es hidrostática. Además, por simplicidad, se supondrá que $p_{atm} = 0$.

Para el fondo, únicamente se establece la condición cinemática.

$$(-ub_x - vb_y + w)|_{z=b} = 0. \quad (2.10)$$

2.4. Ecuaciones de aguas someras

La primera suposición para derivar las ecuaciones de aguas someras es que la componente vertical de la aceleración es despreciable. Esto es

$$\frac{dw}{dt} = w_t + uw_x + vw_y + ww_z = 0. \quad (2.11)$$

Entonces, de la ecuación (2.5d) se concluye que

$$p_z = -\rho g.$$

Integrando verticalmente desde un punto z a la superficie s

$$\begin{aligned} \int_z^s p_z dz &= - \int_z^s \rho g dz, \\ p|_s - p|_z &= -\rho g(s - z), \\ p_{atm} - p(z) &= -\rho g(s - z), \\ p(z) &= \rho g(s - z). \end{aligned}$$

Las dos últimas líneas son consecuencia de la condición dinámica $p(x, y, z, t)|_{z=s} = p_{atm} = 0$. Al derivar este resultado respecto a x y y

$$p_x = \rho g s_x, \quad p_y = \rho g s_y.$$

Notemos que p_x y p_y no depende de z , de donde las componentes de la aceleración del agua, $\frac{du}{dt}$ y $\frac{dv}{dt}$, tampoco son dependientes de z . Por lo tanto, u y v son independientes de z . Así, u_z y v_z son cero. A partir de este análisis las ecuaciones (2.5b) y (2.5c) se reducen a

$$u_t + uu_x + vv_y = -gs_x, \quad (2.12a)$$

$$v_t + uv_x + vv_y = -gs_y. \quad (2.12b)$$

Estas ecuaciones son las ecuaciones de momento y junto con la ecuación de continuidad, (2.5a), forman un sistema con el cual derivamos las ecuaciones de aguas someras.

Integrando la ecuación de continuidad (2.5a) respecto a z desde el fondo hasta la superficie del agua

$$\begin{aligned} \int_b^s (u_x + v_y + w_z) dz &= 0, \\ w|_s - w|_b + \int_b^s (u_x + v_y) dz &= 0. \end{aligned} \quad (2.13)$$

Para conocer los primeros dos términos de la izquierda se usan las condiciones de frontera cinemáticas (2.8) y (2.10), de donde

$$w|_s = (s_t + us_x + vs_y)|_s, \quad w|_b = (ub_x + vb_y)|_b. \quad (2.14)$$

Sustituyendo (2.14) en (2.13),

$$(s_t + us_x + vs_y)|_s - (ub_x + vb_y)|_b + \int_b^s u_x dz + \int_b^s v_y dz = 0.$$

Para calcular las integrales se emplea la regla de Leibniz.

$$\frac{d}{d\xi} \int_{\eta_1(\xi)}^{\eta_2(\xi)} f(\eta, \xi) d\eta = \int_{\eta_1(\xi)}^{\eta_2(\xi)} \frac{df(\eta, \xi)}{d\xi} d\eta + f(\eta_2, \xi) \frac{d\eta_2}{d\xi} - f(\eta_1, \xi) \frac{d\eta_1}{d\xi}.$$

Así, escribimos la ecuación como

$$(s_t + us_x + vs_y)|_s - (ub_x + vb_y)|_b + \frac{\partial}{\partial x} \int_b^s u dz - (us_x)|_s + (ub_x)|_b + \frac{\partial}{\partial y} \int_b^s v dz - (vs_y)|_s + (vb_y)|_b = 0.$$

Simplificando la expresión anterior

$$(s_t)|_s + \frac{\partial}{\partial x} \int_b^s u dz + \frac{\partial}{\partial y} \int_b^s v dz = 0.$$

Recordando que $s = b + h$, que u y v no dependen de z y que b no depende de t , podemos deducir

$$\begin{aligned} s_t + (u(s - b))_x + (v(s - b))_y &= 0 \\ (b + h)_t + (uh)_h + (vh)_y &= 0, \\ h_t + (uh)_x + (vh)_y &= 0. \end{aligned} \tag{2.15}$$

La relación (2.15) es la ecuación de conservación de la masa en forma diferencial.

Para llevar las ecuaciones (2.12) a su forma diferencial procedemos de la manera siguiente. Multiplicamos la ecuación (2.15) por u y la sumamos con la ecuación (2.12a) multiplicada por h , resultando en

$$\begin{aligned} u(h_t + (uh)_x + (vh)_y) + h(u_t + uu_x + vu_y) &= -ghs_x, \\ uh_t + hu_t + u((uh)_x + (vh)_y) + huu_x + hvu_y &= -gh(b + h)_x, \\ (uh)_t + u(hu_x + uh_x + hv_y + vh_y) + huu_x + hvu_y &= -gh(b_x + h_x), \\ (uh)_t + u(2hu_x + uh_x) + gh h_x + (huv_y + uvh_y + hvu_y) &= -ghb_x, \\ (uh)_t + \left(hu^2 + \frac{1}{2}gh^2 \right)_x + (huv)_y &= -ghb_x. \end{aligned} \tag{2.16}$$

Análogamente, se obtiene

$$(vh)_t + (huv)_x + \left(hv^2 + \frac{1}{2}gh^2 \right)_y = -ghb_y. \tag{2.17}$$

Las ecuaciones (2.16) y (2.17) son las ecuaciones de conservación de momento en forma diferencial y junto con la ecuación (2.15) forman el sistema de ecuaciones de aguas poco profundas (2.18).

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} + \frac{\partial \mathbf{g}(\mathbf{u})}{\partial y} = \mathbf{S}(\mathbf{u}), \tag{2.18}$$

donde

$$\mathbf{u} = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \end{pmatrix}, \quad \mathbf{g}(\mathbf{u}) = \begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \end{pmatrix}, \quad \mathbf{S}(\mathbf{u}) = \begin{pmatrix} 0 \\ -ghb_x \\ -ghb_y \end{pmatrix}.$$

Para finalizar, debido a que en este trabajo se consideran las ecuaciones de aguas someras en una dimensión el sistema utilizado con el que trabajamos es

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = \mathbf{S}(\mathbf{u}) \quad (2.19a)$$

con

$$\mathbf{u} = \begin{pmatrix} h \\ hu \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{pmatrix}, \quad \mathbf{S}(\mathbf{u}) = \begin{pmatrix} 0 \\ -ghb_x \end{pmatrix} \quad (2.19b)$$

Otra forma para estas ecuaciones, que considera al caudal o descarga del flujo, $q = hu$, en lugar de la velocidad, u , es decir, $\mathbf{u} = (h, q)^T$, es presentada en el capítulo siguiente. Esta última es la que se usa en los algoritmos desarrollados en esta trabajo.

2.5. Propiedades de las ecuaciones de aguas someras

Se sugiere revisar la definición de invariantes de Riemann, junto con la metodología para calcularlas en el apéndice A antes de leer este apartado. Esto para tener una idea más clara de los conceptos utilizados en los siguientes apartados.

Las invariantes de Riemann son funciones asociadas al problema homogéneo

$$\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} = 0, \quad (2.20a)$$

$$\frac{\partial hu}{\partial t} + \frac{\partial}{\partial x} \left(hu^2 + \frac{1}{2}gh^2 \right) = 0, \quad (2.20b)$$

donde x es la longitud a lo largo del canal, $h > 0$ la profundidad del canal, u la velocidad del flujo y g la aceleración de la gravedad. Éstas pueden ser reescritas en la forma (2.19),

$$\mathbf{u} = \begin{pmatrix} h \\ hu \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} hu \\ \frac{1}{2}gh^2 + hu^2 \end{pmatrix}. \quad (2.21)$$

Luego, la matriz jacobiana será

$$\mathbf{B}(\mathbf{u}) = \begin{pmatrix} \frac{\partial f_1}{\partial h} & \frac{\partial f_1}{\partial hu} \\ \frac{\partial f_2}{\partial h} & \frac{\partial f_2}{\partial hu} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ gh - u^2 & 2u \end{pmatrix}. \quad (2.22)$$

Para hallar los valores propios, $\lambda_k(\mathbf{u})$, con $k = 1, 2$, calculamos el polinomio característico, $p(\lambda) = \lambda^2 - 2u\lambda + (u^2 - gh)$, y buscamos sus raíces.

$$\begin{aligned} \lambda &= \frac{2u \pm \sqrt{4u^2 - 4(u^2 - gh)}}{2} \\ &= u \pm \sqrt{gh}. \end{aligned}$$

De esta manera, los valores propios son

$$\lambda_1(\mathbf{u}) = u - \sqrt{gh}, \quad \lambda_2(\mathbf{u}) = u + \sqrt{gh}. \quad (2.23)$$

Notemos que si $h > 0$, estos valores serán reales y distintos, por lo cual el problema de aguas someras (2.20) es un sistema hiperbólico.

Posteriormente, el vector propio derecho asociado a λ_1 , $\mathbf{s}_1 = (s_1, s_2)^T$ es determinado por

$$(\mathbf{B}(\mathbf{u}) - \lambda_1(\mathbf{u})\mathbf{I}) \mathbf{s}_1 = 0,$$

$$\begin{pmatrix} \sqrt{gh} - u & 1 \\ gh - u^2 & 2u - u + \sqrt{gh} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dado que

$$\begin{pmatrix} \sqrt{gh} - u & 1 \\ gh - u^2 & u + \sqrt{gh} \end{pmatrix} \sim \begin{pmatrix} 1 & \frac{1}{\sqrt{gh} - u} \\ 0 & 0 \end{pmatrix},$$

entonces

$$s_2 = -(\sqrt{gh} - u) s_1,$$

así

$$s_2 = \lambda_1 s_1.$$

Para $s_1 = 1$ se encuentra que un vector propio derecho es

$$\mathbf{s}_1(\mathbf{u}) = (1, \lambda_1)^T. \quad (2.24)$$

Análogamente se obtiene un vector propio derecho \mathbf{s}_2 asociado a λ_2 ,

$$\mathbf{s}_2(\mathbf{u}) = (1, \lambda_2)^T. \quad (2.25)$$

Para encontrar las invariantes de Riemann de este sistema, ver definición A.9, se buscan funciones $w_1(\mathbf{u})$ y $w_2(\mathbf{u})$ que satisfagan la ecuaciones

$$\frac{dw_k}{d\mathbf{u}} \cdot \mathbf{s}_k(\mathbf{u}) = 0, \quad k = 1, 2. \quad (2.26)$$

Para ello, primero necesitamos obtener una matriz de vectores propios izquierdos \mathbf{T} , tal que $\mathbf{TB} = \mathbf{AT}$. Como primer intento para determinar esta matriz, podemos utilizar $\mathbf{T} = \mathbf{S}^{-1}$, donde \mathbf{S} es la matriz cuyas columnas son los de vectores propios derechos \mathbf{s}_1 y \mathbf{s}_2 ,

$$\mathbf{S} = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}.$$

De esta manera,

$$\mathbf{T} = \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} = \frac{1}{2c} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix}.$$

con $c = \sqrt{gh}$. De donde, se deduce que unos vectores propios izquierdos son

$$\mathbf{t}_1 = \frac{1}{2c} (\lambda_2 \quad -1), \quad \mathbf{t}_2 = \frac{1}{2c} (-\lambda_1 \quad 1). \quad (2.27)$$

Luego, por medio de las ecuaciones de Pfaff, (A.31), podemos encontrar la variable $\tilde{\mathbf{u}}$, cuyas componentes serán las invariantes de Riemann,

$$\begin{aligned} d\tilde{\mathbf{u}} &= \mathbf{T}d\mathbf{u} \\ &= \frac{1}{2c} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} \begin{pmatrix} dh \\ d(hu) \end{pmatrix} \\ &= \frac{1}{2c} \begin{pmatrix} \lambda_2 dh - hdu - udh \\ -\lambda_1 dh + udh + hdu \end{pmatrix} \\ &= \frac{1}{2c} \begin{pmatrix} cdh - hdu \\ cdh + hdu \end{pmatrix}. \end{aligned}$$

Es decir,

$$d\tilde{\mathbf{u}} = \begin{pmatrix} \frac{1}{2}dh - \frac{h}{2c}du \\ \frac{1}{2}dh + \frac{h}{2c}du \end{pmatrix}. \quad (2.28)$$

La ecuación (2.28) no es fácil de integrar. Sin embargo, si multiplicamos los vectores propios \mathbf{t}_1 y \mathbf{t}_2 por los factores $-2c/h$ y $2c/h$, respectivamente, obtenemos nuevos vectores propios izquierdos,

$$\mathbf{t}_1 = \frac{1}{h}(-\lambda_2 \ 1), \quad \mathbf{t}_2 = \frac{1}{h}(-\lambda_1 \ 1), \quad (2.29)$$

con los cuales la ecuación de Pfaff asociada a ellos se escribe como

$$d\tilde{\mathbf{u}} = \begin{pmatrix} -\frac{c}{h}dh + du \\ \frac{c}{h}dh + du \end{pmatrix}. \quad (2.30)$$

Observemos que ambas ecuaciones en el sistema (2.30) son EDP exactas y pueden integrarse fácilmente, para dar

$$\tilde{u}_1 = u - 2\sqrt{gh}, \quad \tilde{u}_2 = u + 2\sqrt{gh}. \quad (2.31)$$

Ahora verificaremos que \tilde{u}_1 y \tilde{u}_2 son las invariantes de Riemann para lo cual empleamos la definición (2.26).

Para $\mathbf{s}_1 = (1, \lambda_1)^T$, sea $w_1(\mathbf{u}) = \tilde{u}_2$, así

$$\begin{aligned} \frac{dw_1}{d\mathbf{u}} \cdot \mathbf{s}_1 &= -\frac{hu}{h^2} + \frac{g}{\sqrt{gh}} + \lambda_1 \frac{1}{h} \\ &= -\frac{hu}{h^2} + \frac{g}{\sqrt{gh}} + \frac{u}{h} - \frac{\sqrt{gh}}{h} \\ &= 0. \end{aligned}$$

Similarmente, para $\mathbf{s}_2 = (1, \lambda_2)^T$, definimos $w_2(\mathbf{u}) = \tilde{u}_1$.

Nota: Observemos que $\frac{\partial \tilde{u}_k}{\partial \mathbf{u}} \cdot \mathbf{s}_k \neq 0$, $k = 1, 2$.

2.6. El problema de Riemann para las ecuaciones de aguas someras

El problema de Riemann asociado al sistema de ecuaciones de aguas poco profundas (2.19) es

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0 \quad (2.32a)$$

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_l, & x < 0, \\ \mathbf{u}_r, & x > 0. \end{cases} \quad (2.32b)$$

Se puede demostrar que los vectores propios (2.24) y (2.25) son genuinamente no lineales, ver definición A.8, implicando que la solución del problema (2.32) consta de a lo más tres estados constantes, u_l , u_m y u_r , separados por ondas de rarefacción o de choque. Los posibles patrones de las soluciones se muestran en la figura 2.2.

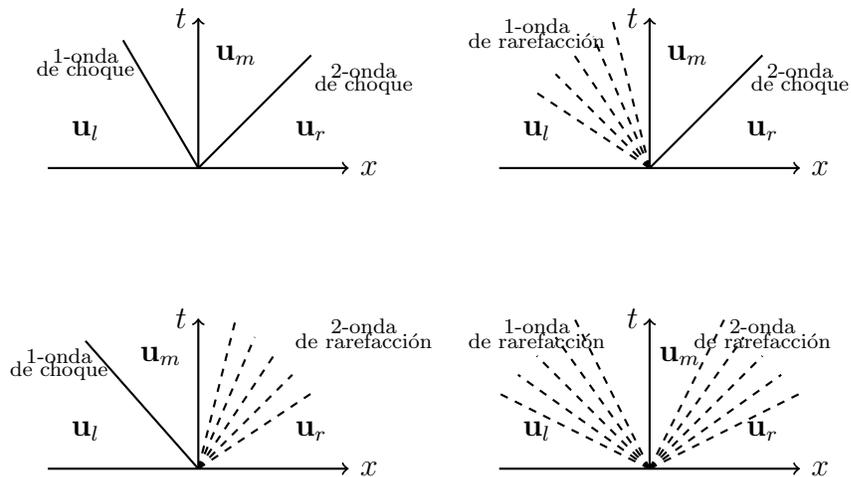


Figura 2.2: Posibles patrones de onda para el problema de Riemann con las ecuaciones de aguas someras. (Las k-ondas de rarefacción son las rectas punteadas, las rectas completas son las k-ondas de choque.)

Se necesita analizar bajo que suposiciones se llega a cada uno de estos escenarios. Para ello, primero suponemos que los estados \mathbf{u}_l y \mathbf{u}_m están separados por una onda de choque, llamada 1-onda de choque y se busca una relación entre \mathbf{u}_l y \mathbf{u}_m por medio de la condición de entropía y la condición de salto de Rankine-Hugoniot. Sea s_1 la velocidad de choque. Para realizar este análisis se definen la siguientes variables:

$$q := hu, \quad v_1 := u - s_1, \quad \mu_1 := hv_1, \quad (2.33)$$

donde v_1 es la velocidad de flujo relativo al choque. Las condiciones de entropía para la 1-onda de choque son

$$s_1 < \lambda_1(\mathbf{u}_l), \quad (2.34a)$$

$$\lambda_1(\mathbf{u}_m) < s_1 < \lambda_2(\mathbf{u}_m), \quad (2.34b)$$

ver definición A.10. Consideremos que $\mathbf{u}_l = (h_l, q_l)^T$ y las definiciones (2.33), entonces

$$\begin{aligned} s_1 &< \lambda_1(\mathbf{u}_l) \\ u_l - v_{1,l} &< u_l - \sqrt{gh_l}, \\ \sqrt{gh_l} &< v_{1,l}. \end{aligned}$$

Análogamente, para $\mathbf{u}_m = (h_m, q_m)^T$ se obtiene

$$\begin{aligned} \lambda_1(\mathbf{u}_m) &< s_1 < \lambda_2(\mathbf{u}_m), \\ u_m - \sqrt{gh_m} &< s_1 < u_m + \sqrt{gh_m}, \\ u_m - \sqrt{gh_m} &< u_m - v_{1,m} < u_m + \sqrt{gh_m}, \\ -\sqrt{gh_m} &< -v_{1,m} < \sqrt{gh_m}, \\ -\sqrt{gh_m} &< v_{1,m} < \sqrt{gh_m}, \\ |v_{1,m}| &< \sqrt{gh_m}. \end{aligned}$$

Sintetizando, se tiene

$$\sqrt{gh_l} < v_{1,l}, \quad (2.35a)$$

$$|v_{1,m}| < \sqrt{gh_m}. \quad (2.35b)$$

Las condiciones de salto de Rankine-Hugoniot, ver ecuación (A.44), del sistema (2.19) se escriben en cada componente como

$$s_1[h]_-^+ = [q]_-^+, \quad (2.36a)$$

$$s_1[q]_-^+ = \left[\frac{q^2}{h} + \frac{gh^2}{2} \right]_-^+, \quad (2.36b)$$

donde $[a]_-^+ = a_m - a_l$.

Al usar las variables v_1 y μ_1 dadas en (2.33) se pueden reescribir estas condiciones como

$$\begin{aligned} s_1[h]_-^+ &= [q]_-^+, \\ s_1(h_m - h_l) &= q_m - q_l, \\ s_1 h_m - s_1 h_l &= q_m - q_l, \\ (u_m - v_{1,m}) h_m - (u_l - v_{1,l}) h_l &= q_m - q_l, \\ q_m - v_{1,m} h_m - q_l + v_{1,l} h_l &= q_m - q_l, \end{aligned}$$

$$\begin{aligned} -v_{1,m}h_m + v_{1,l}h_l &= 0, \\ \mu_{1,m} - \mu_{1,l} &= 0, \\ [\mu_1]_-^+ &= 0. \end{aligned}$$

De donde $\mu_1 = \mu_{1,m} = \mu_{1,l}$. De la segunda condición (2.36b),

$$\begin{aligned} s_1[q]_-^+ &= \left[\frac{q^2}{h} + \frac{gh^2}{2} \right]_-^+, \\ s_1(q_m - q_l) &= \frac{q_m^2}{h_m} + \frac{gh_m^2}{2} - \frac{q_l^2}{h_l} - \frac{gh_l^2}{2}, \\ s_1q_m - s_1q_l &= \frac{q_m^2}{h_m} - \frac{q_l^2}{h_l} + \frac{g}{2}(h_m^2 - h_l^2), \\ \left(\frac{q_m}{h_m} - v_{1,m} \right) q_m - \left(\frac{q_l}{h_l} - v_{1,l} \right) q_l &= \frac{q_m^2}{h_m} - \frac{q_l^2}{h_l} + \frac{g}{2}(h_m^2 - h_l^2), \\ \frac{q_m^2}{h_m} - v_{1,m}q_m - \frac{q_l^2}{h_l} + v_{1,l}q_l &= \frac{q_m^2}{h_m} - \frac{q_l^2}{h_l} + \frac{g}{2}(h_m^2 - h_l^2), \\ -v_{1,m}q_m + v_{1,l}q_l &= \frac{g}{2}(h_m^2 - h_l^2), \\ -\frac{\mu_{1,m}}{h_m}q_m + \frac{\mu_{1,l}}{h_l}q_l &= \frac{g}{2}(h_m^2 - h_l^2), \end{aligned}$$

como $u_k = \frac{q_k}{h_k}$,

$$\begin{aligned} -\mu_{1,m}u_m + s_1\mu_{1,m} - s_1\mu_{1,l} + \mu_{1,l}u_l &= \frac{g}{2}(h_m^2 - h_l^2), \\ -\mu_{1,m}(u_m - s_1) + \mu_{1,l}(u_l - s_1) &= \frac{g}{2}(h_m^2 - h_l^2), \\ -\mu_{1,m}v_{1,m} + \mu_{1,l}v_{1,l} &= \frac{g}{2}(h_m^2 - h_l^2). \end{aligned}$$

Es decir

$$[\mu_1]_-^+ = 0, \tag{2.37a}$$

$$\left[\mu_1 v_1 + \frac{g}{2} h^2 \right]_-^+ = 0. \tag{2.37b}$$

Combinando las condiciones de salto (2.37) y despejando a μ_1 ,

$$\begin{aligned} \left[\mu_1 v_1 + \frac{g}{2} h^2 \right]_-^+ &= 0, \\ \mu_1 \left(\frac{\mu_1}{h_m} - \frac{\mu_1}{h_l} \right) &= -\frac{g}{2}(h_m^2 - h_l^2), \\ \mu_1^2 \left(\frac{h_l - h_m}{h_m h_l} \right) &= \frac{g}{2}(h_l^2 - h_m^2), \end{aligned}$$

$$\mu_1 = \sqrt{\frac{g}{2} (h_l^2 - h_m^2) \frac{h_m h_l}{h_l - h_m}},$$

$$\mu_1 = \sqrt{\frac{g}{2} (h_m + h_l) h_m h_l}.$$

Por otro lado, de la misma ecuación (2.37), pero ahora despejando $v_{1,m} - v_{1,l}$

$$\begin{aligned}\mu_{1,m} v_{1,m} - \mu_{1,l} v_{1,l} &= \frac{g}{2} (h_l^2 - h_m^2), \\ \mu_1 (v_{1,m} - v_{1,l}) &= \frac{g}{2} (h_l^2 - h_m^2), \\ v_{1,m} + s_1 - v_{1,l} - s_1 &= \frac{g}{2\mu_1} (h_l^2 - h_m^2), \\ u_m - u_l &= -\frac{g}{2\mu_1} (h_m^2 - h_l^2).\end{aligned}$$

Resumiendo, se ha llegado a:

$$\mu_1 = \sqrt{\frac{g}{2} (h_m + h_l) h_m h_l}, \quad (2.38a)$$

$$u_m - u_l = -\frac{g}{2\mu_1} (h_m^2 - h_l^2). \quad (2.38b)$$

Sea $z_1 = \frac{h_m}{h_l}$. Para ver como cambia la altura al cruzar la 1-onda de choque se toman la primer condición de salto (2.37) y las desigualdades (2.35), con $v_{1,m} > 0$.

$$\begin{aligned}\mu_{1,m} &= \mu_{1,l}, \\ h_m v_{1,m} &= h_l v_{1,l}, \\ \frac{h_m}{h_l} &= \frac{v_{1,l}}{v_{1,m}}, \\ \frac{h_m}{h_l} &> \sqrt{\frac{gh_l}{gh_m}}, \\ z_1 &> \frac{1}{\sqrt{z_1}}, \\ \frac{h_m}{h_l} &> 1, \\ z_1 &> 1,\end{aligned}$$

esto es, la altura aumenta cuando cruza la 1-onda de choque de izquierda a derecha.

Para analizar el cambio de las velocidades al pasar por la 1-onda de choque se combinan

las ecuaciones en (2.38)

$$\begin{aligned}
u_m - u_l &= -\frac{g}{2\mu_1} (h_m^2 - h_l^2) = -\frac{g}{2} \frac{h_m^2 - h_l^2}{\sqrt{\frac{g}{2} (h_m + h_l) h_m h_l}} \\
&= -\frac{1}{2} \sqrt{2g} \sqrt{h_m - h_l} \sqrt{\frac{h_m}{h_l} - \frac{h_l}{h_m}} \\
&= -\frac{1}{2} \sqrt{2gh_l} \sqrt{\frac{h_m}{h_l}} - 1 \sqrt{\frac{h_m}{h_l} - \frac{h_l}{h_m}} \\
&= -\frac{1}{2} \sqrt{2gh_l} \sqrt{z_1 - 1} \sqrt{z_1 - \frac{1}{z_1}} = -\frac{1}{2} \sqrt{2gh_l} \sqrt{z_1 - 1} \sqrt{\frac{z_1^2 - 1}{z_1}} \\
&= -\frac{1}{2} \sqrt{2gh_l} \sqrt{z_1 - 1} \sqrt{\frac{(z_1 - 1)(z_1 + 1)}{z_1}} = -\frac{1}{\sqrt{2}} \sqrt{gh_l} (z_1 - 1) \sqrt{1 + \frac{1}{z_1}}
\end{aligned}$$

que implica

$$\frac{u_m - u_l}{\sqrt{gh_l}} = -\frac{1}{\sqrt{2}} (z_1 - 1) \sqrt{1 + \frac{1}{z_1}}. \quad (2.39)$$

Dado que $z_1 > 1$, se concluye que la velocidad de flujo decrece cuando pasa la 1-onda de choque de izquierda a derecha.

Ahora, asumimos que los estados \mathbf{u}_l y \mathbf{u}_m están separados por una onda de rarefacción, llamada 1-onda de rarefacción o 1-onda simple. En este caso se tiene que $\lambda_1(\mathbf{u}_l) < \lambda_1(\mathbf{u}_m)$, y la onda de rarefacción consistirá de los rayos $x/t = \lambda_1(\mathbf{u})$ y es acotada por ambos lados por una 1-característica.

$$\begin{aligned}
\lambda_1(\mathbf{u}_m) &> \lambda_1(\mathbf{u}_l), \\
u_m - \sqrt{gh_m} &> u_l - \sqrt{gh_l}, \\
u_m - u_l &> \sqrt{gh_m} - \sqrt{gh_l}.
\end{aligned} \quad (2.40)$$

Para encontrar la solución se ocupa un teorema demostrado en (Godlewski y Raviart, 1996):

Teorema 2.6.1. *Sobre una onda de rarefacción todas las k -invariantes de Riemann son constantes.*

Así, la invariante de Riemann w_1 será constante a través de la 1-onda de rarefacción, dando lugar a

$$\begin{aligned}
w_1(\mathbf{u}) &= w_1(\mathbf{u}_l) = w_1(\mathbf{u}_m) \\
u + 2\sqrt{gh} &= u_l + 2\sqrt{gh_l} = u_m + 2\sqrt{gh_m}, \\
u_m - u_l &= 2(\sqrt{gh_l} - \sqrt{gh_m}),
\end{aligned}$$

$$u_m - u_l = 2\sqrt{gh_l} \left(1 - \frac{\sqrt{gh_m}}{\sqrt{gh_l}} \right).$$

De donde se desprende que

$$\frac{u_m - u_l}{\sqrt{gh_l}} = 2(1 - \sqrt{z_1}). \quad (2.41)$$

Además, de la desigualdad (2.40) se llega a

$$2(1 - \sqrt{z_1}) = \frac{u_m - u_l}{\sqrt{gh_l}} > \frac{\sqrt{gh_m} - \sqrt{gh_l}}{\sqrt{gh_l}} = \sqrt{z_1} - 1.$$

Que implica que $\sqrt{z_1} < 1$. Esto significa que la altura disminuye y la velocidad de flujo aumenta cuando atraviesa la 1-onda de rarefacción de izquierda a derecha.

Todo esto se puede expresar en la siguiente función para la 1-onda que conecta los estados \mathbf{u}_l y \mathbf{u}_m .

$$\frac{u_m - u_l}{\sqrt{gh_l}} = \Gamma(z_1), \quad (2.42a)$$

$$z_1 = \frac{h_m}{h_l}, \quad (2.42b)$$

donde $\Gamma(z)$ está definida por

$$\Gamma(z) = \begin{cases} 2(1 - \sqrt{z}), & \text{si } 0 < z \leq 1, \\ \frac{1}{\sqrt{2}}(1 - z)\sqrt{1 + \frac{1}{z}}, & \text{si } z > 1, \end{cases} \quad (2.43)$$

ver ecuaciones (2.41) y (2.39). Entonces, se tiene que

1. La altura aumenta y la velocidad disminuye al cruzar la 1-onda de choque.
2. La altura disminuye y la velocidad aumenta al cruzar la 1-onda de rarefacción.

Haciendo un análisis similar para los estados \mathbf{u}_r y \mathbf{u}_m , se llega a que

$$\frac{u_m - u_r}{\sqrt{gh_r}} = -\Gamma(z_2), \quad (2.44a)$$

con

$$z_2 = \frac{h_m}{h_r}. \quad (2.44b)$$

Con las propiedades

1. La altura y la velocidad disminuyen al cruzar la 2-onda de choque.

2. La altura y la velocidad aumentan al cruzar la 2-onda de rarefacción.

Para hallar la solución se eliminan los estados intermedios a partir de las ecuaciones (2.42a) y (2.44a) y de las ecuaciones (2.42b) y (2.44b) para obtener:

$$z_1 h_l = z_2 h_r, \quad (2.45a)$$

$$u_l + \sqrt{gh_l} \Gamma(z_1) = u_r - \sqrt{gh_r} \Gamma(z_2). \quad (2.45b)$$

Sean

$$A = \frac{h_r}{h_l}, \quad B = \frac{u_r - u_l}{\sqrt{gh_l}}.$$

Entonces,

$$z_2 = \frac{z_1}{A}$$

y z_1 es raíz de la ecuación

$$G(z_1) := \Gamma(z_1) + \sqrt{A} \Gamma\left(\frac{z_1}{A}\right) - B = 0. \quad (2.46)$$

Es sencillo comprobar que se cumple $G(0) = 2(1 + \sqrt{A}) - B$, $G'(z) < 0$ y $G''(z) > 0$; así la ecuación (2.46) tendrá solución única si $G(0) > 0$ o equivalentemente $u_r - u_l < 2(\sqrt{gh_r} + \sqrt{gh_l})$.

El algoritmo siguiente encuentra la solución al problema de Riemann:

1. Calcular A y B .
2. Encontrar la raíz para la ecuación (2.46). Esto con ayuda de un método iterativo como Newton-Raphson, donde el punto inicial se elige como $G(0) = 2(1 + \sqrt{A}) - B$.
3. Calcular z_2 .
4. Calcular los valores intermedios \mathbf{u}_m .
5. Determinar el patrón que tiene la solución. Es decir, que tipo de ondas son la que separan los estados constantes.

Para determinar cuáles son los patrones que la solución tendrá en términos de A y B se realiza lo siguiente:

Si la 1-onda es una onda de choque se tiene que $z_1 > 1$, de donde

$$\begin{aligned} \Gamma(z_1) &< 0, \\ 0 &< -\Gamma(z_1) = \frac{-u_m + u_l}{\sqrt{gh_l}}, \end{aligned}$$

$$\begin{aligned}
0 < \frac{\sqrt{gh_r}\Gamma(z_2) - u_r + u_l}{\sqrt{gh_l}} &= \sqrt{\frac{gh_r}{gh_l}}\Gamma(z_2) - \frac{u_r - u_l}{\sqrt{gh_l}}, \\
0 < \sqrt{A}\Gamma(z_1/A) - B, \\
B < \sqrt{A}\Gamma(z_1/A)
\end{aligned}$$

En caso contrario, la 1-onda es una onda de rarefacción cuando $B \geq \sqrt{A}\Gamma(z_1/A)$. Del mismo modo, la 2-onda será una onda de choque cuando $B < \Gamma(A)$ y una onda de rarefacción en caso contrario. Así, la solución podrá ser una combinación entre las siguientes 1-onda y 2-onda

1-onda de choque si $B < \sqrt{A}\Gamma\left(\frac{z_1}{A}\right)$:

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_l, & \text{si } x/t < s_1, \\ \mathbf{u}_m, & \text{si } x/t > s_1, \end{cases}$$

con

$$s_1 = u_l - \sqrt{gh_l}\sqrt{\frac{z_1}{2}(1+z_1)}.$$

1-onda de rarefacción si $B \geq \Gamma\left(\frac{z_1}{A}\right)$

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_l, & \text{si } x/t < \lambda_1(\mathbf{u}_l), \\ \left\{ \begin{array}{l} \text{Resolver:} \\ w_1(\mathbf{u}) = w_1(\mathbf{u}_l), \\ \lambda_1(\mathbf{u}) = x/t. \end{array} \right. & \text{si } \lambda_1(\mathbf{u}_l) < x/t < \lambda_1(\mathbf{u}_m), \\ \mathbf{u}_m & \text{si, } x/t > \lambda_1(\mathbf{u}_m). \end{cases}$$

2-onda de choque si $B < \Gamma(A)$

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_m, & \text{si } x/t < s_2, \\ \mathbf{u}_r, & \text{si } x/t > s_2. \end{cases}$$

con

$$s_1 = u_r + \sqrt{gh_r}\sqrt{\frac{z_2}{2}(1+z_2)}.$$

2-onda de rarefacción si $B \geq \Gamma(A)$

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_m, & \text{si } x/t < \lambda_2(\mathbf{u}_m), \\ \left\{ \begin{array}{l} \text{Resolver:} \\ w_2(\mathbf{u}) = w_2(\mathbf{u}_r), \\ \lambda_2(\mathbf{u}) = x/t. \end{array} \right. & \text{si } \lambda_2(\mathbf{u}_m) < x/t < \lambda_2(\mathbf{u}_r), \\ \mathbf{u}_r & \text{si } x/t > \lambda_2(\mathbf{u}_m). \end{cases}$$

Con esto hemos resuelto el problema de Riemann para las ecuaciones de aguas someras.

Capítulo 3

Método de Galerkin discontinuo para el problema de estudio

En este capítulo detallaremos el método de discretización para el espacio en una dimensión empleado en este trabajo. Presentaremos la formulación con Galerkin discontinuo de las ecuaciones de Saint-Venant, las funciones de forma, las transformaciones isoparamétricas, las fórmulas de integración y los flujos numéricos empleados en este trabajo. Un desarrollo más amplio de la teoría de Galerkin discontinuo puede consultarse en (Dolejší y Feistauer, 2015; Hesthaven y Warburton, 2007; Li, 2005).

3.1. Ecuaciones de Saint-Venant

El modelo general que estudiaremos en el resto de este manuscrito está determinado por las ecuaciones de Saint-Venant, dado por

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = \mathbf{S}(\mathbf{u}, x, t), \quad x \in [L, R] = \Omega, \quad t > 0, \quad (3.1a)$$

donde \mathbf{u} es el vector de variables dependientes, $\mathbf{f}(\mathbf{u})$ es la función de flujo de la ecuación y $\mathbf{S}(\mathbf{u}, x, t)$ es el vector de términos fuente, definidos como

$$\mathbf{u} = \begin{pmatrix} h \\ q \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} q \\ \frac{q^2}{h} + \frac{gh^2}{2} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 0 \\ gh(S_0 - S_f) \end{pmatrix}, \quad (3.1b)$$

con h la profundidad del agua, q el caudal o descarga del flujo, g la aceleración de la gravedad, S_0 la pendiente en el fondo del canal y S_f un término que depende de la fricción en la cama del canal expresados por

$$S_0 = -\frac{\partial b(x, t)}{\partial x}, \quad S_f = \frac{m^2 q |q|}{h^{10/3}} \quad (3.1c)$$

con $b(x, t)$ la elevación de la batimetría o cama del canal y m el coeficiente de Manning. Como no consideramos cambio en el fondo del canal, $b(x, t) = b(x)$. Recordemos también que $q = hu$, donde u es la velocidad.

3.2. Discretización del dominio

Sea Ω un dominio acotado en \mathbb{R} con frontera Lipschitz $\partial\Omega$. Además, sea \mathcal{T}_h una partición de la cerradura del dominio, $\bar{\Omega}$, de ne elementos conformes no traslapados, $\Omega_e = [x_e, x_{e+1}]$, con x_e y x_{e+1} los extremos izquierdo y derecho del elemento, respectivamente, esto es

$$\bar{\Omega} = \bigcup_{e=1}^{ne} \Omega_e. \quad (3.2)$$

Estos conjuntos satisfacen que sus interiores son disjuntos dos a dos, $\overset{\circ}{\Omega}_e \cap \overset{\circ}{\Omega}_i = \emptyset$, como se muestra en la figura 3.1.

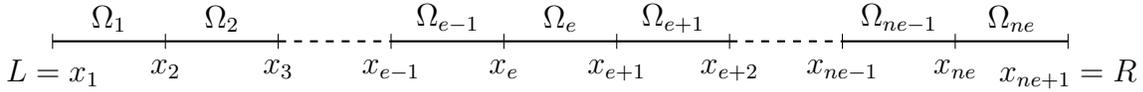


Figura 3.1: Elementos unidimensionales

La frontera del dominio $\partial\Omega$ puede descomponerse como

$$\partial\Omega = \partial\Omega_N \cup \partial\Omega_D,$$

donde $\partial\Omega_N$ es la parte de la frontera con condiciones de frontera Neumann y $\partial\Omega_D$ es la parte de la frontera con condiciones de frontera Dirichlet.

Por su parte, decimos que dos elementos de la triangulación, $\Omega_e, \Omega_i \in \mathcal{T}_h$, son vecinos si $\Omega_e \cap \Omega_i \neq \emptyset$. Además, indicamos que $\Gamma_e \subset \Omega_e$ es una interfaz de Ω_e si es un subconjunto que cumple alguna de las propiedades siguientes:

- $\Gamma_e = \Omega_e \cap \Omega_i$, con Ω_e y Ω_i vecinos, o bien
- $\Gamma_e = \Omega_e \cap \partial\Omega_D$, o bien
- $\Gamma_e = \Omega_e \cap \partial\Omega_N$.

En nuestro caso, las interfaces son simplemente puntos.

Además, un elemento tiene la misma forma que en el método de elemento finito, esto es, $\Omega_e = [x_e, x_{e+1}]$; sin embargo, el esquema de Galerkin discontinuo necesita mantenerse local. Por esta razón duplicamos las variables en los nodos x_e que sean interfaces entre elementos, obteniendo un arreglo de incógnitas, \mathbf{u}_h , que contiene la aproximación a los valores de la

solución \mathbf{u} en cada nodo de la malla. En particular, cuando utilizamos elementos lineales el número de columnas de este arreglo es $2ne$ y tiene la expresión

$$\mathbf{u}_h = (\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{u}_1^2, \mathbf{u}_2^2, \mathbf{u}_1^3, \mathbf{u}_2^3, \dots, \mathbf{u}_1^{ne-1}, \mathbf{u}_2^{ne-1}, \mathbf{u}_1^{ne}, \mathbf{u}_2^{ne}). \quad (3.3)$$

Otra notación para el elemento que usamos a lo largo de este capítulo es $\Omega_e = [x_l^e, x_r^e]$, donde x_l^e y x_r^e son los extremos izquierdo y derecho del elemento, respectivamente. Los valores de las incógnitas sobre estos extremos los denotamos por \mathbf{u}_l^e y \mathbf{u}_r^e , definidos en general por

$$\mathbf{u}_l^e = \lim_{x \rightarrow x_e^+} \mathbf{u}, \quad \mathbf{u}_r^e = \lim_{x \rightarrow x_{e+1}^-} \mathbf{u}, \quad (3.4)$$

En el caso de elementos lineales tendremos que $\Omega_e = [x_1^e, x_2^e]$ y

$$\mathbf{u}_1^e = \lim_{x \rightarrow x_e^+} \mathbf{u}, \quad \mathbf{u}_2^e = \lim_{x \rightarrow x_{e+1}^-} \mathbf{u}, \quad (3.5)$$

como muestra la figura 3.2.

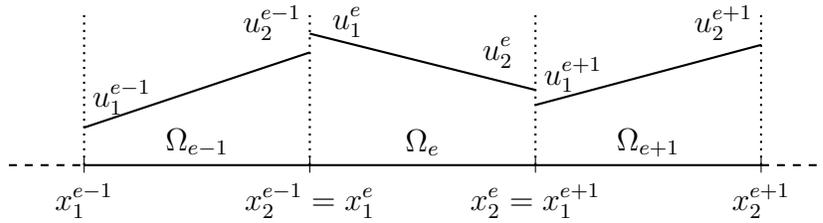


Figura 3.2: Elementos unidimensionales lineales discontinuos

Esto provoca discontinuidades de la solución en las interfaces de los elementos. Para recuperar el valor de la solución global en el nodo x_e se toma el promedio de las soluciones \mathbf{u}_l^e y \mathbf{u}_r^{e-1} . Así, el vector que contiene la solución global aproximada $\tilde{\mathbf{u}}_h$ de \mathbf{u} de tamaño $ne + 1$ se expresa en la ecuación (3.6).

$$\tilde{\mathbf{u}}_h = (\mathbf{u}_h^1, \mathbf{u}_h^2, \mathbf{u}_h^3, \dots, \mathbf{u}_h^{ne-1}, \mathbf{u}_h^{ne}, \mathbf{u}_h^{ne+1}), \quad (3.6a)$$

con

$$\mathbf{u}_h^j = \frac{1}{2} (\mathbf{u}_r^{j-1} + \mathbf{u}_l^j), \quad \text{para } j = 2, 3, \dots, ne. \quad (3.6b)$$

Cabe señalar que esta idea para recuperar la solución global no es única. En (Li, 2005) se muestra que podemos utilizar una combinación ponderada de los valores de la solución definidas sobre las interfaces

$$\mathbf{u}_h^j = \alpha \mathbf{u}_r^{j-1} + (1 - \alpha) \mathbf{u}_l^j.$$

3.3. Definición de espacios y normas

Antes de presentar el método introducimos los siguientes conceptos. Sea $L^2(\Omega)$ el espacio de funciones cuadrado integrables sobre Ω ,

$$L^2(\Omega) = \left\{ u : \int_{\Omega} u^2 dx < \infty, \right\},$$

que tiene asociados el producto interior y la norma global,

$$(u, v)_{\Omega} = \int_{\Omega} uv dx, \quad \|u\|_{\Omega}^2 = (u, u)_{\Omega}, \quad \forall u, v \in L^2(\Omega).$$

Sea $L^2(\Omega_e)$ el espacio de funciones cuadrado integrables sobre Ω_e ,

$$L^2(\Omega_e) = \left\{ u : \int_{\Omega_e} u^2 dx < \infty, \right\},$$

con sus respectivos producto interior local y norma local dados por,

$$(u, v)_{\Omega_e} = \int_{\Omega_e} uv dx, \quad \|u\|_{\Omega_e}^2 = (u, u)_{\Omega_e}, \quad \forall u, v \in L^2(\Omega_e).$$

Dado que el método de Galerkin discontinuo está basado en el uso de aproximaciones discontinuas, definimos el *espacio de Sobolev particionado* sobre una triangulación \mathcal{T}_h , para cualquier $k \in \mathbb{N}$:

$$H^k(\Omega, \mathcal{T}_h) := \{v \in L^2(\Omega) : v|_{\Omega_e} \in H^k(\Omega_e), \forall \Omega_e \in \mathcal{T}_h\}, \quad (3.7)$$

el cual consiste de las funciones cuadrado integrables en Ω cuya restricción en cada elemento de la triangulación pertenece al espacio de Sobolev

$$H^k(\Omega_e) := \{v \in L^2(\Omega_e) : D^{\alpha} v \in L^2(\Omega_e), \alpha \leq k\}.$$

Podemos notar que, en general, las funciones en $H^k(\Omega, \mathcal{T}_h)$ son discontinuas sobre las interfaces de elementos vecinos. Por ello, $H^k(\Omega, \mathcal{T}_h) \not\subset H^k(\Omega)$. Por otro lado, para toda función $v \in H^k(\Omega, \mathcal{T}_h)$ podemos definir la norma

$$\|v\|_{H^k(\Omega, \mathcal{T}_h)}^2 = \sum_{e=1}^{N_e} \|v\|_{H^k(\Omega_e)}^2$$

y la seminorma

$$|v|_{H^k(\Omega, \mathcal{T}_h)}^2 = \sum_{e=1}^{N_e} |v|_{H^k(\Omega_e)}^2$$

3.4. Formulación variacional en espacios de Sobolev particionados

Como se mencionó, $H^1(\Omega, \mathcal{T}_h) \not\subset H^1(\Omega)$, por lo que una formulación variacional clásica sobre todo el dominio Ω no es adecuada debido a las discontinuidades en las interfaces de los elementos. Por ello, en este apartado introducimos la formulación variacional sobre espacios de Sobolev particionados.

Sea $V \subset H^1(\Omega, \mathcal{T}_h)$ un espacio de funciones de prueba en Ω y sea $\mathbf{V} = V \times V$. El *problema variacional* consiste en encontrar una función $\mathbf{u} \in \mathbf{V} \times [0, T_f]$ tal que para cualquier función de prueba $\mathbf{v} \in \mathbf{V}$ se cumpla el sistema

$$\sum_{e=1}^{ne} \int_{\Omega_e} \left(\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} \right) \cdot \mathbf{v} dx = \sum_{e=1}^{ne} \int_{\Omega_e} \mathbf{S}(x, t) \cdot \mathbf{v} dx,$$

separando las ecuaciones

$$\sum_{e=1}^{ne} \left(\int_{\Omega_e} \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{v} dx + \int_{\Omega_e} \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} \cdot \mathbf{v} dx \right) = \sum_{e=1}^{ne} \int_{\Omega_e} \mathbf{S}(x, t) \cdot \mathbf{v} dx, \quad \forall \mathbf{v} \in \mathbf{V}.$$

Aplicando integración por partes al segundo sumando del lado izquierdo

$$\sum_{e=1}^{ne} \left(\int_{\Omega_e} \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{v} dx + (\mathbf{f}(\mathbf{u}) \cdot \mathbf{v}) \Big|_{x_l^e}^{x_r^e} - \int_{\Omega_e} \mathbf{f}(\mathbf{u}) \cdot \frac{\partial \mathbf{v}}{\partial x} dx \right) = \sum_{e=1}^{ne} \int_{\Omega_e} \mathbf{S}(x, t) \cdot \mathbf{v} dx, \quad \forall \mathbf{v} \in \mathbf{V}.$$

Ordenando términos resulta

$$\begin{aligned} \sum_{e=1}^{ne} \left(\int_{\Omega_e} \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{v} dx - \int_{\Omega_e} \mathbf{f}(\mathbf{u}) \cdot \frac{\partial \mathbf{v}}{\partial x} dx \right) \\ = \sum_{e=1}^{ne} \left(\int_{\Omega_e} \mathbf{S}(x, t) \cdot \mathbf{v} dx - (\mathbf{f}(\mathbf{u}) \cdot \mathbf{v}) \Big|_{x_l^e}^{x_r^e} \right), \quad \forall \mathbf{v} \in \mathbf{V}. \end{aligned} \quad (3.9)$$

La ecuación (3.9) es la *formulación variacional* del sistema de ecuaciones (3.1).

3.5. Discretización del problema

El *método de Galerkin discontinuo* puede ser caracterizado como una técnica de elemento finito que utiliza polinomios discontinuos por pedazos. Así, debemos introducir un subespacio de dimensión finita contenido en $H^k(\Omega, \mathcal{T}_h)$, donde buscamos la aproximación a la solución. Con este fin, para \mathcal{T}_h una triangulación de Ω y $p \geq 0$ un entero, definimos el espacio de

funciones polinómicas a trozos como,

$$V_h := \{v_h \in L^2(\Omega) : v_h|_{\Omega_e} \in \mathbb{P}_p(\Omega_e), \forall \Omega_e \in \mathcal{T}_h\}, \quad (3.10)$$

donde $\mathbb{P}_p(\Omega_e)$ es el espacio de polinomios de grado menor o igual a p sobre Ω_e . El número p es el grado de la aproximación polinomial. Notemos que V_h es un espacio de dimensión finita, es un conjunto de funciones discontinuas y $V_h \subset H^k(\Omega, \mathcal{T}_h)$ con $k \in \mathbb{N}$.

Sea $\mathbf{V}_h = V_h \times V_h$. Estamos interesados en encontrar una aproximación $\mathbf{u}_h(t, x) \in \mathbf{V}_h \times [0, T_f]$ a la solución $\mathbf{u}(t, x)$. Para hallarla, reemplazamos \mathbf{u} por \mathbf{u}_h y las funciones de prueba \mathbf{v} por \mathbf{v}_h en la formulación variacional (3.9), obteniendo:

$$\begin{aligned} \sum_{e=1}^{ne} \left(\int_{\Omega_e} \frac{\partial \mathbf{u}_h(t, x)}{\partial t} \cdot \mathbf{v}_h(x) dx - \int_{\Omega_e} \mathbf{f}(\mathbf{u}_h(t, x)) \cdot \frac{\partial \mathbf{v}_h(x)}{\partial x} dx \right) \\ = \sum_{e=1}^{ne} \left(\int_{\Omega_e} \mathbf{S}(x, t) \cdot \mathbf{v}_h(x) dx - (\mathbf{f}(\mathbf{u}_h(t, x)) \cdot \mathbf{v}_h(x)) \Big|_{x_l^e}^{x_r^e} \right) \end{aligned} \quad (3.11)$$

Vemos que en este caso el espacio de funciones de prueba para la formulación discreta (3.11) coincide con el espacio donde buscamos la solución aproximada en x , sin embargo estos espacios se verán modificados dependiendo de las condiciones de frontera del problema.

Definimos para cada $\Omega_e \in \mathcal{T}_h$ un subespacio local de funciones de prueba $\mathbf{V}_h^e = \text{gen} \{ \boldsymbol{\psi}_j^e(\Omega_e) \}_{j=1}^{nne}$. Buscamos la solución local $\mathbf{u}_h^e(x, t) \in \mathbf{V}_h^e \times [0, T_f]$ escrita como una combinación lineal de la base polinomial de \mathbf{V}_h^e :

$$\mathbf{u}_h^e(x, t) = \sum_{j=1}^{nne} \mathbf{u}_j^e(t) \cdot \boldsymbol{\psi}_j^e(x), \quad \forall x \in \Omega_e, \quad (3.12)$$

con $\boldsymbol{\psi}_j^e$ las funciones de base local o funciones de forma generadas por polinomios a trozos sobre el elemento Ω_e , nne el número de nodos por elemento y las \mathbf{u}_j^e están definidas en (3.5). Es importante enfatizar que la solución aproximada de Galerkin discontinuo \mathbf{u}_h , dada por (3.3), restringida a Ω_e es \mathbf{u}_h^e .

Luego, para cada uno de los subespacios \mathbf{V}_h^e , se tiene la formulación variacional local:

$$\begin{aligned} \int_{\Omega_e} \frac{\partial \mathbf{u}_h^e}{\partial t} \cdot \boldsymbol{\psi}_i^e(x) dx - \int_{\Omega_e} \mathbf{f}(\mathbf{u}_h^e) \cdot \frac{\partial \boldsymbol{\psi}_i^e(x)}{\partial x} dx \\ = \int_{\Omega_e} \mathbf{S}(x, t) \cdot \boldsymbol{\psi}_i^e(x) dx - (\mathbf{f}(\mathbf{u}_h^e) \cdot \boldsymbol{\psi}_i^e(x)) \Big|_{x_l^e}^{x_r^e}, \quad 1 \leq i \leq nne. \end{aligned} \quad (3.13)$$

Notemos que la función de flujo \mathbf{f} se debe evaluar en \mathbf{u}_h^e , pero la solución en las interfaces está multivaluada, es decir, en el extremo izquierdo del elemento Ω_e la función de flujo depende de $\{\mathbf{u}_r^{e-1}, \mathbf{u}_l^e\}$ mientras que en el extremo derecho lo hace de $\{\mathbf{u}_r^e, \mathbf{u}_l^{e+1}\}$. Por ello es

necesario determinar qué solución o combinación de las soluciones en las interfaces se debe ocupar. Al flujo evaluado en esta solución se le conoce como *flujo numérico* y es denotado por $\tilde{\mathbf{f}}$. El flujo numérico está vinculado a la ecuación que estemos resolviendo y existen diversas técnicas para calcularlo, llevando a varios esquemas numéricos. Empleando la notación de flujo numérico, el sistema semidiscreto (3.13) se reescribe como

$$\int_{\Omega_e} \left(\frac{\partial \mathbf{u}_h^e}{\partial t} \cdot \boldsymbol{\psi}_i^e(x) - \mathbf{f}(\mathbf{u}_h^e) \cdot \frac{\partial \boldsymbol{\psi}_i^e(x)}{\partial x} \right) dx = \int_{\Omega_e} \mathbf{S} \cdot \boldsymbol{\psi}_i^e(x) dx - \left[\tilde{\mathbf{f}} \cdot \boldsymbol{\psi}_i^e(x) \right]_{x_l^e}^{x_r^e}. \quad (3.14)$$

En el último apartado de este capítulo veremos la obtención de los flujos numéricos utilizados para este trabajo. Ahí observaremos que el flujo numérico enlaza los elementos al determinar de que vecino entra flujo y hacia cual se dirige, logrando un método para encontrar la solución global.

Al sustituir la ecuación (3.12) en (3.14) obtenemos

$$\begin{aligned} \int_{\Omega_e} \left(\frac{\partial}{\partial t} \left(\sum_{j=1}^{nne} \mathbf{u}_j^e(t) \cdot \boldsymbol{\psi}_j^e(x) \right) \cdot \boldsymbol{\psi}_i^e(x) - \mathbf{f}(\mathbf{u}_h^e) \cdot \frac{\partial \boldsymbol{\psi}_i^e(x)}{\partial x} \right) dx \\ = \int_{\Omega_e} \mathbf{S} \cdot \boldsymbol{\psi}_i^e(x) dx - \left[\tilde{\mathbf{f}} \cdot \boldsymbol{\psi}_i^e(x) \right]_{x_l^e}^{x_r^e}, \end{aligned}$$

es decir,

$$\begin{aligned} \sum_{j=1}^{nne} \frac{\partial \mathbf{u}_j^e(t)}{\partial t} \int_{\Omega_e} \boldsymbol{\psi}_j^e(x) \cdot \boldsymbol{\psi}_i^e(x) dx - \int_{\Omega_e} \mathbf{f}(\mathbf{u}_j^e) \cdot \frac{\partial \boldsymbol{\psi}_i^e(x)}{\partial x} dx \\ = \int_{\Omega_e} \mathbf{S} \cdot \boldsymbol{\psi}_i^e(x) dx - \left[\tilde{\mathbf{f}} \cdot \boldsymbol{\psi}_i^e(x) \right]_{x_l^e}^{x_r^e}. \end{aligned}$$

Dado que

$$\begin{aligned} \mathbf{u}_h^e(x, t) &= \sum_{j=1}^{nne} \mathbf{u}_j^e(t) \cdot \boldsymbol{\psi}_j^e(x), \quad \forall x \in \Omega_e, \\ &= \left(\sum_{j=1}^{nne} h_j^e(t) \boldsymbol{\psi}_j^e(x), \sum_{j=1}^{nne} q_j^e(t) \boldsymbol{\psi}_j^e(x) \right)^T, \quad \forall x \in \Omega_e, \end{aligned} \quad (3.15)$$

el sistema semidiscreto se expresa por componentes de la forma siguiente:

$$\sum_{j=1}^{nne} \frac{\partial h_j^e(t)}{\partial t} \int_{\Omega_e} \boldsymbol{\psi}_j^e(x) \cdot \boldsymbol{\psi}_i^e(x) dx - \sum_{j=1}^{nne} q_j^e(t) \int_{\Omega_e} \boldsymbol{\psi}_j^e(x) \cdot \frac{\partial \boldsymbol{\psi}_i^e(x)}{\partial x} dx = - \left[\tilde{f}_1 \boldsymbol{\psi}_i^e(x) \right]_{x_l^e}^{x_r^e}$$

$$\begin{aligned} \sum_{j=1}^{nne} \frac{\partial q_j^e(t)}{\partial t} \int_{\Omega_e} \psi_j^e(x) \cdot \psi_i^e(x) dx - \int_{\Omega_e} \left(\frac{\left(\sum_{j=1}^{nne} q_j^e(t) \psi_j^e(x) \right)^2}{\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x)} + \frac{1}{2} g \left(\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x) \right)^2 \right) \frac{\partial \psi_i^e(x)}{\partial x} dx \\ = \int_{\Omega_e} g \left(\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x) \right) (S_0 - S_f) \psi_i^e(x) dx - \left[\tilde{f}_2 \psi_i^e(x) \right]_{x_i^e}^{x_r^e}, \end{aligned}$$

donde \tilde{f}_1 y \tilde{f}_2 son las componentes del flujo numérico $\tilde{\mathbf{f}}$.

Las integrales que involucran al flujo y al término fuente de las ecuaciones son aproximadas por medio de la regla del trapecio debido a que usamos elementos lineales. Así llegamos a los sistemas (3.16a) y (3.16b) de nne ecuaciones algebraicas lineales por nne incógnitas cada uno ha resolverse para cada Ω_e ,

$$\mathbf{M}^e \frac{d\mathbf{h}^e}{dt} - \mathbf{N}^e \mathbf{q}^e = - \left[\tilde{f}_1 \boldsymbol{\psi}^e \right]_{x_i^e}^{x_r^e}, \quad (3.16a)$$

$$\mathbf{M}^e \frac{d\mathbf{q}^e}{dt} - \mathbf{A}^e = \mathbf{B}^e - \left[\tilde{f}_2 \boldsymbol{\psi}^e \right]_{x_i^e}^{x_r^e}, \quad (3.16b)$$

donde las entradas de la matrices \mathbf{M}^e y \mathbf{N}^e están dadas por

$$\mathbf{M}_{ij}^e = \int_{\Omega_e} \psi_i^e(x) \psi_j^e(x) dx, \quad \mathbf{N}_{ij}^e = \int_{\Omega_e} \psi_j^e(x) \frac{\partial \psi_i^e(x)}{\partial x} dx;$$

\mathbf{h}^e y \mathbf{q}^e son los vectores de las soluciones locales y $\boldsymbol{\psi}^e$ es el vector de funciones de forma o de base local,

$$\mathbf{h}^e = [h_1^e, \dots, h_{nne}^e]^T, \quad \mathbf{q}^e = [q_1^e, \dots, q_{nne}^e]^T, \quad \boldsymbol{\psi}^e = [\psi_1^e(x), \dots, \psi_{nne}^e(x)]^T$$

Por otro lado, las entradas de los vectores \mathbf{A}^e y \mathbf{B}^e tienen la forma

$$\mathbf{A}_i^e = \int_{\Omega_e} \left(\frac{\left(\sum_{j=1}^{nne} q_j^e(t) \psi_j^e(x) \right)^2}{\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x)} + \frac{1}{2} g \left(\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x) \right)^2 \right) \frac{\partial \psi_i^e(x)}{\partial x} dx, \quad 1 \leq i \leq nne$$

$$\mathbf{B}_i^e = g \int_{\Omega_e} \left(\sum_{j=1}^{nne} h_j^e(t) \psi_j^e(x) \right) (S_0(x, t) - S_f(x, t)) \psi_i^e(x) dx, \quad 1 \leq i \leq nne$$

El esquema (3.16) es *el método de Galerkin discontinuo* para las ecuaciones de aguas someras.

3.6. Estimación del error para el método de Galerkin discontinuo

En (Godlewski y Raviart, 1996; Hesthaven y Warburton, 2007) se hace un análisis del error para el esquema de Galerkin discontinuo. En ellos se establece que una cota para el error está dada por

$$\|u - u_h\|_{H^k(\Omega, \mathcal{T}_h)} \leq C_1 h^{\mu-1} |u|_{H^\mu(\Omega)},$$

donde, $u \in H^s(\Omega)$, $\mu = \min(p + 1, s)$, h es el tamaño de la malla, p es el grado de la aproximación y C_1 es independiente de h y de u . Así, si $s > p + 1$ entonces, una cota para el error es

$$\|u - u_h\|_{H^k(\Omega, \mathcal{T}_h)} \leq C_1 h^p |u|_{H^{p+1}(\Omega)}.$$

Así, el error para el método de Galerkin discontinuo es de orden $O(h^p)$.

3.7. Funciones de forma

Pasamos ahora a la descripción de las funciones de forma, las cuales definimos como aquella que interpola la solución sobre cada elemento de la triangulación. Estas funciones serán polinomios a trozos. Además, están determinadas por los nne puntos de malla locales y su grado es $nne - 1$.

En particular trabajamos con el conjunto $V_h = \{v \in L^2(\Omega) : v|_{\Omega_e} \in \mathbb{P}_1(\Omega_e), \forall \Omega_e \in \mathcal{T}_h\}$ y las funciones de forma aplicadas en cada V_h^e son funciones lineales dadas por

$$\psi_1^e(x) = \frac{x - x_2^e}{x_1^e - x_2^e}, \quad \psi_2^e(x) = \frac{x - x_1^e}{x_2^e - x_1^e}, \quad (3.17)$$

con la propiedad

$$\psi_i^e(x_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \quad (3.18)$$

como muestra la figura 3.3.

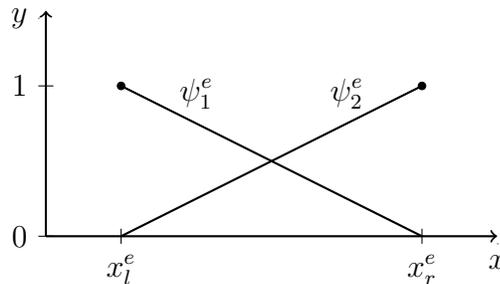


Figura 3.3: Funciones de forma lineales unidimensionales

Notemos que $\boldsymbol{\psi}_i^e = (\psi_i^e, \psi_i^e)^T$. De este modo, la matriz \mathbf{M}^e en (3.16) se escribe como:

$$\begin{aligned} \mathbf{M}^e &= \begin{bmatrix} \int_{\Omega_e} \psi_1^e(x) \psi_1^e(x) dx & \int_{\Omega_e} \psi_1^e(x) \psi_2^e(x) dx \\ \int_{\Omega_e} \psi_2^e(x) \psi_1^e(x) dx & \int_{\Omega_e} \psi_2^e(x) \psi_2^e(x) dx \end{bmatrix} \\ &= \begin{bmatrix} \int_{\Omega_e} \left(\frac{x - x_2^e}{x_1^e - x_2^e} \right)^2 dx & \int_{\Omega_e} \left(\frac{x - x_2^e}{x_1^e - x_2^e} \right) \left(\frac{x - x_1^e}{x_2^e - x_1^e} \right) dx \\ \int_{\Omega_e} \left(\frac{x - x_1^e}{x_2^e - x_1^e} \right) \left(\frac{x - x_2^e}{x_1^e - x_2^e} \right) dx & \int_{\Omega_e} \left(\frac{x - x_1^e}{x_2^e - x_1^e} \right)^2 dx \end{bmatrix} \end{aligned}$$

Aunque las integrales involucradas son sencillas, éstas dependen de los extremos de los elementos, implicando que deben calcularse elemento por elemento, aumentando ligeramente la complejidad. En el siguiente apartado introducimos las transformaciones isoparamétricas, que permiten simplificar los cálculos de estas integrales.

3.8. Transformaciones isoparamétricas

En lugar de trabajar con una base local para cada elemento, (3.17), podemos trabajar con un elemento de referencia $\hat{\Omega}$, simplificando el trabajo. Este elemento es independiente de la malla y tiene su propio sistema de coordenadas ξ . El elemento de referencia se extiende de $\xi = -1$ a $\xi = 1$ y tiene el mismo número de nodos que el elemento Ω_e . De este modo, cada elemento Ω_e puede verse como un mapeo de $\hat{\Omega}$ a Ω_e , figura 3.4, mediante la transformación

$$T_e : \hat{\Omega} \rightarrow \Omega_e, \quad x(\xi) = \frac{1}{2}(x_2^e - x_1^e)\xi + \frac{1}{2}(x_1^e + x_2^e), \quad (3.19)$$

donde x_1^e y x_2^e son los extremos del elemento Ω_e y $\xi \in \hat{\Omega}$. Podemos observar que esta transformación cumple $x(-1) = x_1^e$ y $x(1) = x_2^e$. Además denotamos al tamaño del elemento de referencia como $\Delta\xi = \xi_2 - \xi_1 = 2$.

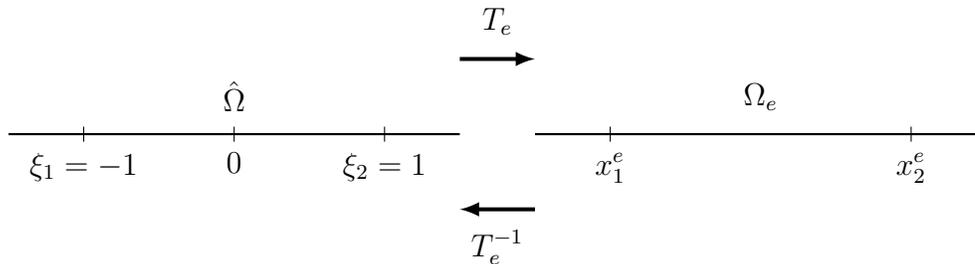


Figura 3.4: Transformada isoparamétrica y su inversa

Una vez definida la transformación, puede crearse una base de referencia $\{\hat{\psi}_1(\xi), \hat{\psi}_2(\xi)\}$ que cumpla con la propiedad (3.18) en $\hat{\Omega}$. Con esto en cuenta, se utiliza (3.19) para mapear $\hat{\psi}_1(\xi)$ y $\hat{\psi}_2(\xi)$ en $\hat{\psi}_1^e(x)$ y $\hat{\psi}_2^e(x)$ respectivamente, que satisfagan

$$\psi_1^e(x) = \psi_1^e(x(\xi)) = \hat{\psi}_1(\xi), \quad \psi_2^e(x) = \psi_2^e(x(\xi)) = \hat{\psi}_2(\xi). \quad (3.20)$$

Luego, para el elemento de referencia $\hat{\Omega}$ la base de referencia está determinada por las funciones

$$\hat{\psi}_1(\xi) = \frac{1-\xi}{2}, \quad \hat{\psi}_2(\xi) = \frac{1+\xi}{2} \quad (3.21)$$

3.9. Integración en el espacio

Con el fin de facilitar la evaluación de las integrales en las entradas de la matriz \mathbf{M}^e , operamos en el elemento de referencia $\hat{\Omega}$ por medio de las funciones de forma definidas en él y el método de cambio de variable. Para esto, primero calculamos el jacobiano de la transformación, resultando en

$$J_e = \frac{dx}{d\xi} = \frac{x_2^e - x_1^e}{2} = \frac{\Delta x^e}{2}.$$

Como $J_e \neq 0$,

$$J_e^{-1} = \frac{d\xi}{dx} = \frac{2}{\Delta x^e}.$$

Para obtener las derivadas de $\psi_i^e(x)$ respecto a x aplicamos la regla de la cadena, así

$$\frac{d\psi_i^e(x)}{dx} = \frac{d\hat{\psi}_i(\xi)}{d\xi} \frac{d\xi}{dx} = J_e^{-1} \frac{d\hat{\psi}_i(\xi)}{d\xi} = \frac{2}{\Delta x^e} \frac{d\hat{\psi}_i(\xi)}{d\xi}.$$

Además

$$\frac{\partial \hat{\psi}_1(\xi)}{\partial \xi} = \frac{-1}{2}, \quad \frac{\partial \hat{\psi}_2(\xi)}{\partial \xi} = \frac{1}{2}.$$

Sustituyendo llegamos a la expresión

$$\frac{d\psi_i^e(x)}{dx} = \frac{(-1)^i}{\Delta x^e}, \quad i = 1, 2, \quad (3.22)$$

luego, la entrada M_{ij}^e se reescribe como

$$M_{ij}^e = \int_{x_i^e}^{x_j^e} \psi_i^e(x) \psi_j^e(x) dx$$

$$\begin{aligned}
&= \int_{-1}^1 \hat{\psi}_i(\xi) \hat{\psi}_j(\xi) J_e d\xi \\
&= \frac{\Delta x^e}{2} \int_{-1}^1 \hat{\psi}_i(\xi) \hat{\psi}_j(\xi) d\xi.
\end{aligned}$$

Finalmente, sustituyendo y calculando, se obtiene

$$\begin{aligned}
\mathbf{M}^e &= \frac{\Delta x^e}{2} \begin{bmatrix} \int_{-1}^1 \hat{\psi}_1(\xi) \hat{\psi}_1(\xi) d\xi & \int_{-1}^1 \hat{\psi}_1(\xi) \hat{\psi}_2(\xi) d\xi \\ \int_{-1}^1 \hat{\psi}_2(\xi) \hat{\psi}_1(\xi) d\xi & \int_{-1}^1 \hat{\psi}_2(\xi) \hat{\psi}_2(\xi) d\xi \end{bmatrix} \\
&= \frac{\Delta x^e}{8} \begin{bmatrix} \int_{-1}^1 (1-\xi)^2 d\xi & \int_{-1}^1 (1-\xi)(1+\xi) d\xi \\ \int_{-1}^1 (1+\xi)(1-\xi) d\xi & \int_{-1}^1 (1+\xi)^2 d\xi \end{bmatrix} \\
&= \frac{\Delta x^e}{8} \begin{bmatrix} \frac{8}{3} & \frac{4}{3} \\ \frac{4}{3} & \frac{8}{3} \end{bmatrix} = \frac{\Delta x^e}{24} \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix}
\end{aligned}$$

Notemos que cuando la malla es uniforme, las matrices elementales \mathbf{M}^e son iguales. Similarmente, aplicando la transformación (3.19) y la ecuación (3.22) para llevar las entradas de la matriz \mathbf{N}^e al elemento de referencia, obtenemos

$$\mathbf{N}_{ij}^e = \frac{(-1)^i}{2} \int_{-1}^1 \hat{\psi}_j(\xi) d\xi.$$

Entonces la matriz \mathbf{N}^e es

$$\begin{aligned}
\mathbf{N}^e &= \frac{1}{2} \begin{bmatrix} -\int_{-1}^1 \hat{\psi}_1(\xi) d\xi & -\int_{-1}^1 \hat{\psi}_2(\xi) d\xi \\ \int_{-1}^1 \hat{\psi}_1(\xi) d\xi & \int_{-1}^1 \hat{\psi}_2(\xi) d\xi \end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix} -\int_{-1}^1 (1-\xi) d\xi & -\int_{-1}^1 (1+\xi) d\xi \\ \int_{-1}^1 (1-\xi) d\xi & \int_{-1}^1 (1+\xi) d\xi \end{bmatrix}
\end{aligned}$$

$$= \frac{1}{4} \begin{bmatrix} -2 & -2 \\ 2 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

En general para integrar una función $g(x)$ sobre Ω_e , primero llevamos las coordenadas x a las del elemento de referencia, ξ , por medio de la transformación isoparamétrica (3.19).

$$\begin{aligned} \int_{x_1^e}^{x_2^e} g(x) dx &= \int_{-1}^1 g(x(\xi)) \frac{dx}{d\xi} d\xi \\ &= \int_{-1}^1 \hat{g}(\xi) J_e d\xi. \end{aligned} \quad (3.23)$$

Así, una entrada del vector \mathbf{A}^e puede calcularse de la forma siguiente:

$$\begin{aligned} \mathbf{A}_i^e &= \int_{\Omega_e} \left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \psi_j^e(x) \right)^2}{\sum_{j=1}^2 h_j^e(t) \psi_j^e(x)} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \psi_j^e(x) \right)^2 \right) \frac{\partial \psi_i^e(x)}{\partial x} dx, \\ &= \int_{-1}^1 \left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \psi_j^e(x(\xi)) \right)^2}{\sum_{j=1}^2 h_j^e(t) \psi_j^e(x(\xi))} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \psi_j^e(x(\xi)) \right)^2 \right) \frac{(-1)^i}{\Delta x^e} J_e d\xi, \\ &= \int_{-1}^1 \left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \hat{\psi}_j(\xi) \right)^2}{\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(\xi)} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(\xi) \right)^2 \right) \frac{(-1)^i}{\Delta x^e} \frac{\Delta x^e}{2} d\xi, \\ &= \int_{-1}^1 \left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \hat{\psi}_j(\xi) \right)^2}{\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(\xi)} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(\xi) \right)^2 \right) \frac{(-1)^i}{2} d\xi, \end{aligned}$$

al aplicar la regla del trapecio resulta

$$\begin{aligned} \mathbf{A}_i^e &\approx \frac{(-1)^i}{2} \left[\left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \hat{\psi}_j(-1) \right)^2}{\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(-1)} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(-1) \right)^2 \right) \right. \\ &\quad \left. + \left(\frac{\left(\sum_{j=1}^2 q_j^e(t) \hat{\psi}_j(1) \right)^2}{\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(1)} + \frac{1}{2} g \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(1) \right)^2 \right) \right] \frac{\Delta \xi}{2}, \end{aligned}$$

$$\begin{aligned} &\approx \frac{(-1)^i}{2} \left[\left(\frac{(q_1^e(t))^2}{h_1^e(t)} + \frac{1}{2}g(h_1^e(t))^2 \right) + \left(\frac{(q_2^e(t))^2}{h_2^e(t)} + \frac{1}{2}g(h_2^e(t))^2 \right) \right], \\ &\approx \frac{(-1)^i}{2} \sum_{j=1}^2 \left(\frac{(q_j^e)^2}{h_j^e} + \frac{1}{2}g(h_j^e)^2 \right). \end{aligned}$$

Finalmente, una coordenada del vector \mathbf{B}^e se puede escribir como

$$\begin{aligned} \mathbf{B}_i^e &= g \int_{\Omega_e} \left(\sum_{j=1}^2 h_j^e(t) \psi_j^e(x) \right) (S_0(x, t) - S_f(x, t)) \psi_i^e(x) dx, \\ &= g \int_{-1}^1 \left(\sum_{j=1}^2 h_j^e(t) \psi_j^e(x(\xi)) \right) (S_0(x(\xi), t) - S_f(x(\xi), t)) \psi_i^e(x(\xi)) J_e d\xi, \\ &= g \int_{-1}^1 \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(\xi) \right) (\hat{S}_0(\xi, t) - \hat{S}_f(\xi, t)) \hat{\psi}_i(\xi) \frac{\Delta x^e}{2} d\xi, \end{aligned}$$

con la regla del trapecio,

$$\begin{aligned} &\approx g \frac{\Delta x^e}{2} \left(\left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(-1) \right) (\hat{S}_0(-1, t) - \hat{S}_f(-1, t)) \hat{\psi}_i(-1) \right. \\ &\quad \left. + \left(\sum_{j=1}^2 h_j^e(t) \hat{\psi}_j(1) \right) (\hat{S}_0(1, t) - \hat{S}_f(1, t)) \hat{\psi}_i(1) \right) \frac{\Delta \xi}{2}, \\ &= g \frac{\Delta x^e}{2} \left(h_1^e(t) (\hat{S}_0(-1, t) - \hat{S}_f(-1, t)) \hat{\psi}_i(-1) + h_2^e(t) (\hat{S}_0(1, t) - \hat{S}_f(1, t)) \hat{\psi}_i(1) \right). \end{aligned}$$

En resumen, se obtienen los vectores

$$\mathbf{A}^e \approx \mathbf{A}_h^e = \frac{1}{2} \sum_{i=1}^2 \left(\frac{(q_i^e)^2}{h_i^e} + \frac{1}{2}g(h_i^e)^2 \right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad (3.24a)$$

$$\mathbf{B}^e \approx \mathbf{B}_h^e = \frac{g\Delta x^e}{2} \begin{bmatrix} h_1^e (S_0^e(-1) - S_f^e(-1)) \\ h_2^e (S_0^e(1) - S_f^e(1)) \end{bmatrix} \quad (3.24b)$$

Con esto, hemos discretizado por completo la parte espacial. Llevando el sistema (3.16) al método discreto

$$\frac{d\mathbf{h}^e}{dt} = (\mathbf{M}^e)^{-1} \left(\mathbf{N}^e \mathbf{q}^e - \begin{bmatrix} \tilde{f}_1 \boldsymbol{\psi}^e \\ \mathbf{x}_l^e \end{bmatrix} \right) \quad (3.25)$$

$$\frac{d\mathbf{q}^e}{dt} = (\mathbf{M}^e)^{-1} \left(\mathbf{A}_h^e + \mathbf{B}_h^e - \left[\tilde{f}_2 \boldsymbol{\psi}^e \right]_{x_l^e}^{x_r^e} \right) \quad (3.26)$$

3.10. Flujos numéricos

El flujo numérico está ligado al método de volumen finito. Por esta razón, partimos de este punto de vista para dar las ideas del flujo numérico. En primer lugar, definimos el esquema de Godunov, el cual sirve como base para introducir el flujo numérico de Roe. Posteriormente, se presenta el flujo de Roe para las ecuaciones de aguas someras. Además exponemos los flujos numéricos HLL y de Lax-Friedrichs.

3.10.1. Esquema de Godunov

Tomemos el problema de valor inicial para la función $\mathbf{u}(x, t)$,

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0, \quad \mathbf{u}(x, 0) = \mathbf{v}_0(x), \quad x \in \mathbb{R}, \quad t > 0, \quad (3.27)$$

con $\mathbf{f}(\mathbf{u})$ la función de flujo que generalmente es no lineal. Supongamos que el problema (3.27) es un sistema hiperbólico lineal, esto es, que la matriz jacobiana $\mathbf{B}(\mathbf{u}) := \mathbf{f}'(\mathbf{u})$ tiene m valores propios reales, $\lambda_1(\mathbf{u}) \leq \lambda_2(\mathbf{u}) \leq \dots \leq \lambda_m(\mathbf{u})$, y m vectores propios derechos linealmente independientes, $\mathbf{s}_1(\mathbf{u}), \mathbf{s}_2(\mathbf{u}), \dots, \mathbf{s}_m(\mathbf{u})$.

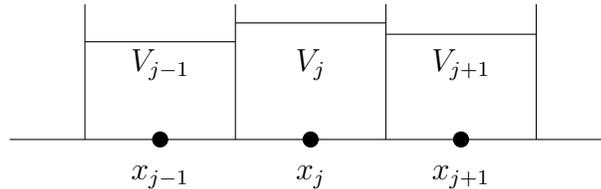


Figura 3.5: Solución numérica constante a trozos en el tiempo t^n

Definimos un volumen de control V_j como sigue:

$$V_j := \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}} \right), \quad x_{j+\frac{1}{2}} = \frac{1}{2}(x_j + x_{j+1}), \quad x_j = j\Delta x, \quad j = 0, \pm 1, \pm 2, \dots, \quad (3.28)$$

sobre el cual la solución es constante, como se muestra en la figura 3.5. La solución numérica \mathbf{u}_j^n sobre el volumen de control V_j se interpreta como una aproximación del valor promedio de la solución sobre el volumen de control V_j en el paso de tiempo t^n . De esta forma, podemos asociar a cada \mathbf{u}_j^n la función vectorial $\check{\mathbf{u}}(x, t)$ de la manera siguiente:

$$\frac{\partial \check{\mathbf{u}}}{\partial t} + \frac{\partial \mathbf{f}(\check{\mathbf{u}})}{\partial x} = \mathbf{0}, \quad x \in \mathbb{R}, \quad t > t^n, \quad (3.29a)$$

$$\check{\mathbf{u}}(x, t^n) = \mathbf{u}_j^n, \quad x \in V_j. \quad (3.29b)$$

El esquema de Godunov está basado en la formulación integral del problema de valor inicial (3.29). Integrando esta ecuación sobre el volumen de control V_j y en el intervalo $[t^n, t^{n+1}]$ llegamos a

$$\begin{aligned} \int_{V_j} \check{\mathbf{u}}(x, t^{n+1}) dx - \int_{V_j} \check{\mathbf{u}}(x, t^n) dx \\ + \int_{t^n}^{t^{n+1}} \mathbf{f}(\check{\mathbf{u}}(x_{j+\frac{1}{2}}, t)) dt - \int_{t^n}^{t^{n+1}} \mathbf{f}(\check{\mathbf{u}}(x_{j-\frac{1}{2}}, t)) dt = \mathbf{0}. \end{aligned} \quad (3.30)$$

Si definimos el flujo numérico como

$$\tilde{\mathbf{f}}(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n) = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{f}(\check{\mathbf{u}}(x_{j+\frac{1}{2}}, t)) dt, \quad (3.31)$$

y el vector \mathbf{u}_j^{n+1} como el vector de valores promedio de $\check{\mathbf{u}}(x, t^{n+1})$ sobre el volumen de control V_j

$$\mathbf{u}_j^{n+1} := \frac{1}{\Delta x} \int_{V_j} \check{\mathbf{u}}(\mathbf{x}, t^{n+1}) dx.$$

entonces la expresión (3.30) se puede reescribir, obteniendo

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\Delta t}{\Delta x} \left(\tilde{\mathbf{f}}(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n) - \tilde{\mathbf{f}}(\mathbf{u}_{j-1}^n, \mathbf{u}_j^n) \right). \quad (3.32)$$

En la ecuación (3.31) hemos escrito de forma explícita que el flujo numérico en la frontera $x_{j+\frac{1}{2}}$ del elemento es una función que depende de los vectores \mathbf{u}_j^n y \mathbf{u}_{j+1}^n . La cuestión ahora es como calcular el valor de los flujos numéricos, sabiendo que se desconocen los valores del integrando.

Para calcular el flujo numérico $\tilde{\mathbf{f}}(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n)$, resolvemos un problema de Riemann en la frontera $x_{j+\frac{1}{2}}$, que consiste de la ecuación (3.29a) y la condición inicial

$$\check{\mathbf{u}}(x, t^n) = \begin{cases} \mathbf{u}_j^n & \text{si } x < x_{j+\frac{1}{2}}, \\ \mathbf{u}_{j+1}^n & \text{si } x > x_{j+\frac{1}{2}}. \end{cases}$$

La solución a este problema es una función de similaridad de la forma

$$\mathbf{u}(x, t) = \mathbf{u}_R(\eta; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n), \quad \eta := \frac{x - x_{j+\frac{1}{2}}}{t - t^n},$$

donde la función \mathbf{u}_R depende de los parámetros η y los vectores \mathbf{u}_j^n y \mathbf{u}_{j+1}^n , ver apéndice A. Vemos que cuando $x = x_{j+\frac{1}{2}}$, $\eta = 0$. Además, para el sistema de Godunov se supone

que esta solución no es influenciada por los problemas de Riemann vecinos, implicando que $\tilde{\mathbf{u}}(x_{j+\frac{1}{2}}, t) = \mathbf{u}_R(0; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n) = C \in \mathbb{R}^m$ en el intervalo $t^n \leq t \leq t^{n+1}$, ver (Mattheij et al., 2005). Al sustituir en la ecuación (3.31) e integrar, el flujo numérico se puede calcular, dando como resultado

$$\tilde{\mathbf{f}}(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n) = \mathbf{f}(\mathbf{u}_R(0; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n)). \quad (3.33)$$

De este análisis podemos concluir que para encontrar los valores de los flujos numéricos en cada una de las interfaces tenemos que resolver problemas de Riemann locales, sin embargo, para problemas no lineales esto puede complicarse considerablemente, como veremos a continuación.

Flujo de Godunov para el caso lineal

Para calcular el flujo numérico de Godunov en el caso lineal se considera el problema de Riemann

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{B} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{0}, \quad x \in \mathbb{R}, \quad t > 0 \quad (3.34a)$$

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_l & \text{si } x < 0, \\ \mathbf{u}_r & \text{si } x > 0. \end{cases} \quad (3.34b)$$

Como los vectores propios de \mathbf{B} son linealmente independientes, los vectores \mathbf{u}_l y \mathbf{u}_r pueden escribirse como

$$\mathbf{u}_l = \sum_{k=1}^m \alpha_k \mathbf{s}_k, \quad \mathbf{u}_r = \sum_{k=1}^m \beta_k \mathbf{s}_k, \quad \alpha, \beta \in \mathbb{R}.$$

Además, en el apéndice A se establece que la solución de similaridad está dada por:

$$\mathbf{u}_R(x/t; \mathbf{u}_l, \mathbf{u}_r) = \sum_{\lambda_k < x/t} \beta_k \mathbf{s}_k + \sum_{\lambda_k > x/t} \alpha_k \mathbf{s}_k.$$

La solución \mathbf{u}_R es una solución constante a trozos debido a que la discontinuidad inicial en $x = 0$ se propaga a lo largo de las características. Las partes constantes de ella están separadas por las características. La solución a lo largo de $x/t = 0$ es

$$\mathbf{u}_R(0; \mathbf{u}_l, \mathbf{u}_r) = \sum_{\lambda_k < 0} \beta_k \mathbf{s}_k + \sum_{\lambda_k > 0} \alpha_k \mathbf{s}_k.$$

Reagrupando términos, se puede llegar a expresiones alternativas

$$\mathbf{u}_R(0; \mathbf{u}_l, \mathbf{u}_r) = \mathbf{u}_l + \sum_{\lambda_k < 0} (\beta_k - \alpha_k) \mathbf{s}_k, \quad (3.35)$$

$$\mathbf{u}_R(0; \mathbf{u}_l, \mathbf{u}_r) = \mathbf{u}_r - \sum_{\lambda_k > 0} (\beta_k - \alpha_k) \mathbf{s}_k. \quad (3.36)$$

Finalmente, utilizando estas expresiones y partiendo de que $\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{B}\mathbf{u}_R(0; \mathbf{u}_l, \mathbf{u}_r)$, podemos obtener el flujo numérico de Godunov

$$\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{B}\mathbf{u}_l + \sum_{\lambda_k < 0} (\beta_k - \alpha_k) \lambda_k \mathbf{s}_k, \quad (3.37)$$

$$\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{B}\mathbf{u}_r - \sum_{\lambda_k > 0} (\beta_k - \alpha_k) \lambda_k \mathbf{s}_k. \quad (3.38)$$

3.10.2. Solvers de Riemann

El esquema de Godunov requiere resolver problemas de Riemann locales en las interfaces de los volúmenes de control para calcular los flujos numéricos. Para los casos no lineales esto resulta ser complicado y costoso. Por ello se emplean *solvers* del problema de Riemann, donde el flujo numérico es aproximado. Algunos de estos flujos fueron propuestos por Roe, Harten, Lax, y van Leer. En los apartados siguientes presentamos su obtención.

Flujo numérico de Roe

Una forma sencilla para encontrar el flujo numérico es remplazar el problema de Riemann no lineal por una aproximación lineal. Este método se debe a Roe.

Como en el esquema de Godunov, el esquema de Roe se escribe en la forma siguiente:

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\Delta t}{\Delta x} \left(\tilde{\mathbf{f}}_R(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n) - \tilde{\mathbf{f}}_R(\mathbf{u}_{j-1}^n, \mathbf{u}_j^n) \right), \quad (3.39)$$

donde el flujo numérico de Roe, $\tilde{\mathbf{f}}_R(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n)$ es una aproximación al flujo de Godunov. La forma en que se calcula el flujo de Roe se describe a continuación. Primero se deriva la expresión integral para el flujo numérico de Godunov del problema de Riemann original. Después, se reemplaza el problema de Riemann por una linealización y nuevamente se determinan formas integrales para el flujo numérico de Godunov correspondiente. Finalmente, combinando estos flujos, se calcula el flujo de Roe.

Sea el problema de Riemann para un sistema de ecuaciones

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0, \quad (3.40a)$$

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_l & \text{si } x < 0, \\ \mathbf{u}_r & \text{si } x > 0. \end{cases} \quad (3.40b)$$

En la figura (3.6) se muestra un esquema de una solución de similaridad $\mathbf{u}_R(x/t; \mathbf{u}_l, \mathbf{u}_r)$, donde s_l y s_r son las velocidades de propagación mínima y máxima respectivamente involucradas en el problema de Riemann. La solución de similaridad \mathbf{u}_R se mantiene en la zona en forma de cuña en el centro y está acotada por los estados constantes \mathbf{u}_l y \mathbf{u}_r .

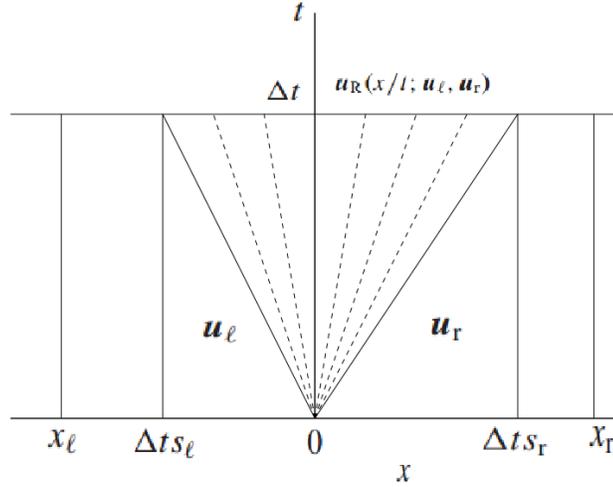


Figura 3.6: Esquema de una solución de similaridad

Al integrar la ecuación (3.40) sobre el volumen de control izquierdo $[x_l, 0] \times [0, \Delta t]$ para algún paso de tiempo $\Delta t > 0$ y tomando en cuenta que la solución es constante en el tiempo fuera de la zona en forma de cuña, se llega a

$$\begin{aligned} & \int_{x_l}^0 \int_0^{\Delta t} \frac{\partial \mathbf{u}}{\partial t} dt dx + \int_0^{\Delta t} \int_{x_l}^0 \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} dx dt = 0, \\ & \int_{x_l}^0 [\mathbf{u}(x, \Delta t) - \mathbf{u}(x, 0)] dx + \int_0^{\Delta t} [\mathbf{f}(\mathbf{u}(0, t)) - \mathbf{f}(\mathbf{u}(x_l, t))] dt = 0, \\ & \int_{x_l}^0 \mathbf{u}(x, \Delta t) dx - \mathbf{u}_l(0 - x_l) + \int_0^{\Delta t} \mathbf{f}(\mathbf{u}(0, t)) dt - \int_0^{\Delta t} \mathbf{f}(\mathbf{u}(x_l, t)) dt = 0, \\ & \int_{x_l}^0 \mathbf{u}(x, \Delta t) dx + x_l \mathbf{u}_l + \int_0^{\Delta t} \mathbf{f}(\mathbf{u}(0, t)) dt - \mathbf{f}(\mathbf{u}(x_l))(\Delta t - 0) = 0, \\ & \tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_l) - \frac{1}{\Delta t} \left(\int_{x_l}^0 \mathbf{u}(x, \Delta t) dx + x_l \mathbf{u}_l \right), \end{aligned} \quad (3.41)$$

donde $\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_R(0; \mathbf{u}_l, \mathbf{u}_r))$ es el flujo de Godunov, ver ecuación (3.33).

Análogamente, al integrar (3.40) en $[0, x_r] \times [0, \Delta t]$ produce

$$\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_r) + \frac{1}{\Delta t} \left(\int_0^{x_r} \mathbf{u}(x, \Delta t) dx - x_r \mathbf{u}_r \right). \quad (3.42)$$

De esta manera conseguimos dos formas integrales para el flujo de Godunov. La igualdad de los lados derechos en ambas expresiones puede confirmarse al integrar (3.40) en $[x_l, x_r] \times [0, \Delta t]$.

El siguiente paso para calcular el flujo de Roe está relacionado con la solución del problema de Riemann

$$\frac{\partial \tilde{\mathbf{u}}}{\partial t} + \tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r) \frac{\partial \tilde{\mathbf{u}}}{\partial x} = \mathbf{0}, \quad (3.43a)$$

$$\tilde{\mathbf{u}}(x, 0) = \begin{cases} \mathbf{u}_l & \text{si } x < 0, \\ \mathbf{u}_r & \text{si } x > 0, \end{cases} \quad (3.43b)$$

con $\tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)$ la matriz de Roe constante que solo depende de los estados iniciales \mathbf{u}_l y \mathbf{u}_r . Claramente este problema es lineal con función de flujo $\hat{\mathbf{f}}(\mathbf{u}) = \tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r) \mathbf{u}$. El flujo $\hat{\mathbf{f}}(\mathbf{u})$ y la matriz $\tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)$ serán aproximaciones al flujo $\mathbf{f}(\mathbf{u})$ en (3.40) y a la matriz jacobiana $\mathbf{B} = \frac{\partial \mathbf{f}}{\partial \mathbf{u}}$. La matriz debe cumplir las propiedades siguientes:

1. $\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l) = \tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)(\mathbf{u}_r - \mathbf{u}_l)$. (Condición de salto de Rankine-Hugoniot (A.19).)
2. $\tilde{\mathbf{B}}(\mathbf{u}, \mathbf{u}) = \mathbf{B}(\mathbf{u})$. (Consistencia de la matriz)
3. $\tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)$ tiene m valores propios $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_m$; asociados a m vectores propios linealmente independientes $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_m$. (El sistema (3.43) también es hiperbólico.)

Para encontrar el flujo numérico de Godunov del problema (3.43) se procede de manera similar a lo realizado para el problema (3.40):

$$\tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) = \hat{\mathbf{f}}(\mathbf{u}_l) - \frac{1}{\Delta t} \left(\int_{x_l}^0 \tilde{\mathbf{u}}(x, \Delta t) dx + x_l \mathbf{u}_l \right), \quad (3.44)$$

$$\tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) = \hat{\mathbf{f}}(\mathbf{u}_r) + \frac{1}{\Delta t} \left(\int_0^{x_r} \tilde{\mathbf{u}}(x, \Delta t) dx - x_r \mathbf{u}_r \right). \quad (3.45)$$

Tomando la diferencia entre (3.41) y (3.44)

$$\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) - \tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_l) - \hat{\mathbf{f}}(\mathbf{u}_l) - \frac{1}{\Delta t} \int_{x_l}^0 (\mathbf{u}(x, \Delta t) - \tilde{\mathbf{u}}(x, \Delta t)) dx,$$

y considerando el valor de la integral pequeño, se obtiene una expresión para el flujo numérico de Roe $\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r)$:

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) + \mathbf{f}(\mathbf{u}_l) - \hat{\mathbf{f}}(\mathbf{u}_l). \quad (3.46)$$

Del mismo modo, trabajando con (3.42) y (3.44) se deduce otra expresión para el flujo numérico de Roe

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) + \mathbf{f}(\mathbf{u}_r) - \hat{\mathbf{f}}(\mathbf{u}_r). \quad (3.47)$$

Lo que falta por hacer es encontrar una expresión adecuada para $\tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r)$. Para esto, se usa la solución de similaridad del problema de Riemann (3.43) en $x/t = 0$ escribiéndola como en las ecuaciones (3.35) y (3.36),

$$\begin{aligned} \tilde{\mathbf{u}}_R(0; \mathbf{u}_l, \mathbf{u}_r) &= \mathbf{u}_l + \sum_{\tilde{\lambda}_k < 0} \gamma_k \tilde{\mathbf{s}}_k, \\ \tilde{\mathbf{u}}_R(0; \mathbf{u}_l, \mathbf{u}_r) &= \mathbf{u}_r - \sum_{\tilde{\lambda}_k > 0} \gamma_k \tilde{\mathbf{s}}_k, \end{aligned}$$

entonces

$$\mathbf{u}_r - \mathbf{u}_l = \sum_{k=1}^m \gamma_k \tilde{\mathbf{s}}_k.$$

Con estas relaciones es posible calcular $\tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r)$, ecuación (3.37)

$$\begin{aligned} \tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) &= \hat{\mathbf{f}}(\mathbf{u}_l) + \sum_{\tilde{\lambda}_k < 0} \gamma_k \tilde{\lambda}_k \tilde{\mathbf{s}}_k, \\ \tilde{\mathbf{F}}(\mathbf{u}_l, \mathbf{u}_r) &= \hat{\mathbf{f}}(\mathbf{u}_r) - \sum_{\tilde{\lambda}_k > 0} \gamma_k \tilde{\lambda}_k \tilde{\mathbf{s}}_k. \end{aligned}$$

Sustituyendo en (3.46) o (3.47) produce dos expresiones para el flujo de Roe

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_l) + \sum_{\tilde{\lambda}_k < 0} \gamma_k \tilde{\lambda}_k \tilde{\mathbf{s}}_k, \quad (3.48a)$$

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_r) - \sum_{\tilde{\lambda}_k > 0} \gamma_k \tilde{\lambda}_k \tilde{\mathbf{s}}_k. \quad (3.48b)$$

Tomando el promedio de estas expresiones da lugar a otra representación del flujo numérico de Roe

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \frac{1}{2} (\mathbf{f}(\mathbf{u}_l) + \mathbf{f}(\mathbf{u}_r)) - \frac{1}{2} \left(\sum_{k=1}^m \gamma_k |\tilde{\lambda}_k| \tilde{\mathbf{s}}_k \right). \quad (3.49)$$

En conclusión, para calcular el flujo de Roe, se necesitan los siguientes pasos:

1. Calcular la matriz de Roe $\tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)$ a partir de $\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l) = \tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)(\mathbf{u}_r - \mathbf{u}_l)$.
2. Encontrar los valores y vectores propios de $\tilde{\mathbf{B}}(\mathbf{u}, \mathbf{u})$.

3. Buscar los coeficientes γ_k .
4. Hallar el flujo numérico a partir de cualquiera de las expresiones en (3.48).

A continuación se aplica esto para encontrar el flujo de Roe para las ecuaciones de aguas someras

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = \mathbf{S}(x, t), \quad x \in [L, R] = \Omega, \quad t > 0, \quad (3.50)$$

donde

$$\mathbf{u} = (h, q)^T, \quad \mathbf{f}(\mathbf{u}) = \left(q, \frac{q^2}{h} + \frac{gh^2}{2} \right)^T, \quad \mathbf{S} = (0, gh(S_0 - S_f))^T.$$

Primero se busca la matriz $\tilde{\mathbf{B}}$ de la relación

$$\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l) = \tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r)(\mathbf{u}_r - \mathbf{u}_l).$$

Se introduce el vector de parámetros \mathbf{z} , de modo que las diferencias $\mathbf{u}_r - \mathbf{u}_l$ y $\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l)$ puedan representarse en términos de $\mathbf{z}_r - \mathbf{z}_l$. Esto de la siguiente forma

$$\mathbf{u}_r - \mathbf{u}_l = \tilde{\mathbf{P}}(\mathbf{u}_l, \mathbf{u}_r)(\mathbf{z}_r - \mathbf{z}_l), \quad (3.51a)$$

$$\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l) = \tilde{\mathbf{Q}}(\mathbf{u}_l, \mathbf{u}_r)(\mathbf{z}_r - \mathbf{z}_l), \quad (3.51b)$$

donde $\tilde{\mathbf{P}}(\mathbf{u}_l, \mathbf{u}_l)$ y $\tilde{\mathbf{Q}}(\mathbf{u}_l, \mathbf{u}_r)$ son matrices de 2×2 que dependen de los estados \mathbf{u}_l y \mathbf{u}_r . Si se supone que la matriz $\tilde{\mathbf{P}}(\mathbf{u}_l, \mathbf{u}_r)$ es no singular se llega a la relación

$$\tilde{\mathbf{B}}(\mathbf{u}_l, \mathbf{u}_r) = \tilde{\mathbf{Q}}(\mathbf{u}_l, \mathbf{u}_r)\tilde{\mathbf{P}}^{-1}(\mathbf{u}_l, \mathbf{u}_r). \quad (3.52)$$

Definiendo el vector de parámetros \mathbf{z} como

$$\mathbf{z} := \frac{1}{\sqrt{h}}\mathbf{u} \quad \text{o} \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \frac{1}{\sqrt{h}} \begin{pmatrix} h \\ q \end{pmatrix}, \quad (3.53)$$

es posible expresar el vector de estados \mathbf{u} y el vector de flujos $\mathbf{f}(\mathbf{u})$ como

$$\mathbf{u} = \begin{pmatrix} z_1^2 \\ z_1 z_2 \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} z_1 z_2 \\ z_2^2 + \frac{1}{2}gz_1^4 \end{pmatrix}. \quad (3.54)$$

Al calcular las diferencias $\mathbf{u}_r - \mathbf{u}_l$ y $\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l)$ resulta

$$\begin{aligned} h_r - h_l &= 2\bar{z}_1(z_{1,r} - z_{1,l}), \\ q_r - q_l &= \bar{z}_2(z_{1,r} - z_{1,l}) + \bar{z}_1(z_{2,r} - z_{2,l}), \end{aligned}$$

$$\begin{aligned}\frac{1}{2}(gh_r^2 - gh_l^2) &= 2\bar{\varphi}\bar{z}_1(z_{1,r} - z_{1,l}), \\ \frac{q_r^2}{h_r} - \frac{q_l^2}{h_l} &= 2\bar{z}_2(z_{2,r} - z_{2,l}).\end{aligned}$$

donde $\varphi = gh$ y $\bar{v} = \frac{1}{2}(v_l + v_r)$ representa el valor promedio de la variable v , Así, las matrices $\tilde{\mathbf{P}}$ y $\tilde{\mathbf{Q}}$ son

$$\tilde{\mathbf{P}} = \begin{bmatrix} 2\bar{z}_1 & 0 \\ \bar{z}_2 & \bar{z}_1 \end{bmatrix}, \quad \tilde{\mathbf{Q}} = \begin{bmatrix} \bar{z}_2 & \bar{z}_1 \\ 2\bar{\varphi}\bar{z}_1 & 2\bar{z}_2 \end{bmatrix}.$$

Claramente la matriz $\tilde{\mathbf{P}}$ es no singular, pues su determinante es $2\bar{z}_1^2 \neq 0$. Luego, a partir de (3.52), se llega a

$$\tilde{\mathbf{B}} = \begin{bmatrix} 0 & 1 \\ \bar{\varphi} - \left(\frac{\bar{z}_2}{\bar{z}_1}\right)^2 & \frac{2\bar{z}_2}{\bar{z}_1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \bar{\varphi} - \tilde{u}^2 & 2\tilde{u} \end{bmatrix}, \quad (3.55)$$

donde \tilde{u} es la velocidad media ponderada, definida por

$$\tilde{u} = \frac{\bar{z}_2}{\bar{z}_1} = \frac{\left(\frac{q}{\sqrt{h}}\right)_l + \left(\frac{q}{\sqrt{h}}\right)_r}{\sqrt{h_l} + \sqrt{h_r}}.$$

Notemos que la matriz de Roe $\tilde{\mathbf{B}}$ es la matriz jacobiana de (3.50), con φ remplazado por $\bar{\varphi}$ y la velocidad de flujo $u = \frac{q}{h}$ remplazada por \tilde{u} . De esta manera se concluye que los valores y vectores propios de $\tilde{\mathbf{B}}$ son

$$\tilde{\lambda}_1(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u} - \tilde{c}, \quad \tilde{\lambda}_2(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u} + \tilde{c}, \quad \tilde{c} = \sqrt{\bar{\varphi}},$$

$$\tilde{\mathbf{s}}_1 = \begin{pmatrix} 1 \\ \tilde{u} - \tilde{c} \end{pmatrix}, \quad \tilde{\mathbf{s}}_2 = \begin{pmatrix} 1 \\ \tilde{u} + \tilde{c} \end{pmatrix}.$$

A partir de estos y de la relación

$$\mathbf{u}_r - \mathbf{u}_l = \sum_{k=1}^m \gamma_k \tilde{\mathbf{s}}_k$$

se calculan los coeficientes γ_k , dando lugar a

$$\begin{aligned}\gamma_1 &= \frac{1}{2\tilde{c}} \left(\tilde{\lambda}_2(h_r - h_l) - (q_r - q_l) \right), \\ \gamma_2 &= \frac{1}{2\tilde{c}} \left(-\tilde{\lambda}_1(h_r - h_l) + (q_r - q_l) \right).\end{aligned}$$

Así, al evaluar en cualquiera de las expresiones de (3.48) se deben distinguir los tres casos siguientes: $\tilde{\lambda}_1 > 0$, $\tilde{\lambda}_1 < 0 < \tilde{\lambda}_2$ y $\tilde{\lambda}_2 < 0$. Para el primer caso, al tomar la expresión (3.48a) se concluye que $\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_l)$. De manera análoga, para el tercer caso y tomando la relación (3.48b), $\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \mathbf{f}(\mathbf{u}_r)$. Finalmente, para el segundo caso, cualquiera de las expresiones puede ser utilizada. Luego, el flujo numérico de Roe para las ecuaciones de aguas poco profundas es

$$\tilde{\mathbf{f}}_R(\mathbf{u}_l, \mathbf{u}_r) = \begin{cases} \mathbf{f}(\mathbf{u}_l) & \text{si } \tilde{\lambda}_1 > 0, \\ \mathbf{f}(\mathbf{u}_l) + \gamma_1 \tilde{\lambda}_1 \tilde{\mathbf{s}}_1 & \text{si } \tilde{\lambda}_1 < 0 < \tilde{\lambda}_2, \\ \mathbf{f}(\mathbf{u}_r) & \text{si } \tilde{\lambda}_2 < 0. \end{cases} \quad (3.56)$$

Flujo numérico HLL

El flujo de Roe es apropiado para muchos problemas. Sin embargo, para otros más puede ser demasiado simple, llevando a la aparición de oscilaciones espurias. Para resolver esto se propone que el problema de Riemann consista de tres estados constantes separados por dos ondas S_l y S_r , con $S_l < S_r$, figura 3.7. Este fue propuesto por Harten, Lax y Van Leer en (Harten et al., 1983).

Tanto la onda más lenta, S_l , como la más rápida, S_r , se suponen conocidas. Y en el caso no trivial $S_l < 0 < S_r$, la solución se obtiene integrando sobre el volumen de control.

Integrando la ley de conservación en el volumen de control se escribe

$$\iint (\mathbf{u}_t + \mathbf{f}_x) dx dt = \mathbf{0}.$$

Aplicando el teorema de Green

$$\int_{\Gamma} (\mathbf{u} dx - \mathbf{f} dt) = \mathbf{0}.$$

Al integrar esta ecuación en el volumen de control izquierdo $[TS_l, 0] \times [0, T]$ se obtiene

$$(\tilde{\mathbf{u}} - \mathbf{u}_l)(0 - TS_l) + T(\tilde{\mathbf{f}} - \mathbf{f}_l) = \mathbf{0}.$$

Al hacerlo en el derecho $[0, TS_r] \times [0, T]$, se encuentra

$$(\tilde{\mathbf{u}} - \mathbf{u}_r)(TS_r - 0) + T(\mathbf{f}_r - \tilde{\mathbf{f}}) = \mathbf{0}.$$

donde los valores \mathbf{f}_r , \mathbf{f}_l y $\tilde{\mathbf{f}}$ son los flujos a través de la frontera del elemento.

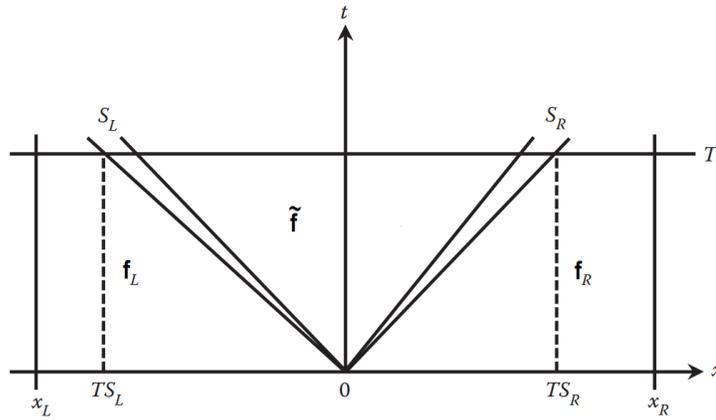


Figura 3.7: Configuración de dos ondas

Despejando $\tilde{\mathbf{u}}$ de ambas ecuaciones

$$\tilde{\mathbf{u}} = \mathbf{u}_l + \frac{\tilde{\mathbf{f}} - \mathbf{f}_l}{S_l},$$

$$\tilde{\mathbf{u}} = \mathbf{u}_r - \frac{\mathbf{f}_r - \tilde{\mathbf{f}}}{S_r}.$$

Luego

$$\mathbf{u}_l + \frac{\tilde{\mathbf{f}} - \mathbf{f}_l}{S_l} = \mathbf{u}_r - \frac{\mathbf{f}_r - \tilde{\mathbf{f}}}{S_r},$$

$$\frac{\mathbf{f}_r - \tilde{\mathbf{f}}}{S_r} + \frac{\tilde{\mathbf{f}} - \mathbf{f}_l}{S_l} = \mathbf{u}_r - \mathbf{u}_l,$$

$$\frac{S_l(\mathbf{f}_r - \tilde{\mathbf{f}}) + S_r(\tilde{\mathbf{f}} - \mathbf{f}_l)}{S_l S_r} = \mathbf{u}_r - \mathbf{u}_l,$$

$$\tilde{\mathbf{f}}(\mathbf{u}_l, \mathbf{u}_r) = \frac{(\mathbf{u}_r - \mathbf{u}_l)(S_l S_r) + S_r \mathbf{f}_l - S_l \mathbf{f}_r}{S_r - S_l}.$$

Siguiendo estas ideas, el flujo HLL es definido como

$$\tilde{\mathbf{f}}_{HLL}(\mathbf{u}_l, \mathbf{u}_r) = \begin{cases} \mathbf{f}(\mathbf{u}_l) & \text{si } S_l > 0 \\ \frac{(\mathbf{u}_r - \mathbf{u}_l)(S_l S_r) + S_r \mathbf{f}_l - S_l \mathbf{f}_r}{S_r - S_l} & \text{si } S_l < 0 < S_r \\ \mathbf{f}(\mathbf{u}_r) & \text{si } S_r < 0 \end{cases} \quad (3.57)$$

Claramente, si $S_l > 0$ la información se propaga hacia la derecha y basta con un esquema tipo “upwind”. De manera similar cuando $S_r < 0$.

Falta estimar los valores de S_l y S_r . Para ello, Harten, Lax y van Leer proponen que las velocidades S_l y S_r sean calculadas como sigue

$$S_l = \min(u_l - \sqrt{gh_l}, u_r - \sqrt{gh_r}) = \min(\lambda_1(\mathbf{u}_l), \lambda_1(\mathbf{u}_r)),$$

$$S_r = \max(u_l + \sqrt{gh_l}, u_r + \sqrt{gh_r}) = \max(\lambda_2(\mathbf{u}_l), \lambda_2(\mathbf{u}_r)),$$

con λ_1 y λ_2 los valores propios de las ecuaciones de aguas someras.

Por otro lado, en (Fraccarollo y Toro, 1995) se sugiere que

$$S_l = \min(\lambda_1(\mathbf{u}_l), u^* - c^*), \quad S_r = \max(\lambda_2(\mathbf{u}_r), u^* + c^*), \quad (3.58)$$

con

$$u^* = \frac{1}{2}(u_l + u_r) + \sqrt{gh_l} - \sqrt{gh_r}, \quad c^* = \frac{1}{2}(\sqrt{gh_l} + \sqrt{gh_r}) + \frac{1}{4}(u_l - u_r).$$

Nos referiremos al flujo numérico con los valores de S_l y S_r dados por la ecuación (3.58) como el flujo HLL-FT.

Flujo numérico de Lax-Friedrichs

El último de los flujos que utilizamos en este trabajo es aquel desarrollado por P. Lax y K. O. Friedrichs. Este método está basado en diferencias finitas, y puede ser descrito como un método centrado en el espacio y hacia adelante en el tiempo que introduce un término de viscosidad igual a $1/2$. Este flujo es descrito por la siguiente expresión

$$\tilde{\mathbf{f}}_{LF}(\mathbf{u}_l, \mathbf{u}_r) = \frac{1}{2} \left(\tilde{\mathbf{f}}(\mathbf{u}_l + \tilde{\mathbf{f}}(\mathbf{u}_r)) - S_{max}(\mathbf{u}_r - \mathbf{u}_l) \right), \quad (3.59)$$

donde

$$S_{max} = \max(|u_l| + \sqrt{gh_l}, |u_r| + \sqrt{gh_r}).$$

Capítulo 4

Integración en el tiempo

Una vez resuelta la parte espacial del problema, podemos escribir el sistema a resolver como una ecuación diferencial ordinaria dependiente del tiempo. Para efectuar la integración en el tiempo utilizamos los métodos de Euler hacia adelante y Runge-Kutta de segundo grado, los cuales se reescriben para que sean estables, ya sea que satisfagan la propiedad de disminuir la variación total (TVD) o que mantengan la variación total acotada en las medias (TVBM). También empleamos limitadores de pendiente. A continuación se detallan estas técnicas.

4.1. Esquemas numéricos para integrar el tiempo

4.1.1. Método de Euler

El método de Euler es la técnica de aproximación más simple para resolver problemas con valores iniciales. Su simplicidad es de ayuda para ejemplificar las técnicas con que se desarrollan algunos métodos más avanzados. Este método tiene como objetivo encontrar una aproximación de un problema con valores iniciales, cuya forma es

$$\frac{du}{dt} = L(t, u), \quad t \in [t^0, T] \quad (4.1)$$

con L un operador y con una condición inicial

$$u(x, t^0) = u^0(x).$$

El método de Euler hacia adelante consiste en dividir el intervalo $[t^0, T]$ en N subintervalos de longitud

$$\Delta t = \frac{T - t^0}{N},$$

generando un conjunto discreto de $N + 1$ puntos en el intervalo $[t^0, T]$. Para cualquiera de estos puntos se cumple

$$t^n = t^0 + n\Delta t, \quad 1 \leq n \leq N$$

De este modo, podemos aproximar la primer derivada como

$$\frac{du}{dt}(t^n) \approx \frac{u^{n+1} - u^n}{\Delta t},$$

con $u^n = u(x, t^n)$. Al sustituir lo anterior en la ecuación (4.1) se encuentra un valor aproximado para la función u en el tiempo t^{n+1} a partir del valor en el tiempo t^n :

$$u^{n+1} = u^n + \Delta t L(t^n, u^n). \quad (4.2)$$

El método de Euler es un esquema de primer orden $O(\Delta t)$.

4.1.2. Método de Runge-Kutta de segundo orden

Por su parte, el método de Runge-Kutta de segundo orden consiste en tomar la forma generalizada del operador L

$$u^{n+1} = u^n + \Delta t L(t^n, u^n, \Delta t), \quad (4.3)$$

que ahora depende del incremento Δt y

$$L(u^n, t^n, \Delta t) = a_1 k_1 + a_2 k_2,$$

con a_1 y a_2 constantes y

$$\begin{aligned} k_1 &= L(t^n, u^n), \\ k_2 &= L(t^n + p_1 \Delta t, u^n + q_{11} k_1 \Delta t). \end{aligned}$$

Los coeficientes se encuentran del sistema de ecuaciones siguiente obtenido al igualar las relaciones anteriores con la serie de Taylor de segundo orden:

$$\begin{aligned} a_1 + a_2 &= 1, \\ a_2 p_1 &= .5, \\ a_2 q_{11} &= .5, \end{aligned}$$

si a_2 es el parámetro libre llegamos a la expresión siguiente:

$$a_1 = 1 - a_2,$$

$$p_1 = q_{11} = \frac{1}{2a_2}.$$

Así, existe una infinidad de posibles métodos de RK de segundo orden. En este trabajo se usa $a_2 = .5$, implicando que $a_1 = .5$, $p_1 = q_{11} = 1$, llamado método de Heun con un solo corrector y se escribe como

$$u^{n+1} = u^n + \frac{\Delta t}{2}(k_1 + k_2), \quad (4.4)$$

con

$$k_1 = L(t^n, u^n), \quad k_2 = L(t^n + \Delta t, u^n + k_1 \Delta t).$$

Este esquema es de segundo orden $O(\Delta t^2)$.

En (Chapra y Canale, 2003) puede consultarse el desarrollo de estos coeficientes y algunos otros ejemplos de métodos de RK de segundo orden y superiores.

Se debe mencionar que ambos métodos son condicionalmente estables. Esto es, la convergencia del método depende de un parámetro, conocido como número de Courant-Friedrich-Lewy (CFL), que depende de la velocidad v , el tamaño de la malla Δx y el paso de tiempo Δt . Este número debe estar acotado para que no aparezcan falsas oscilaciones en la solución.

4.2. Esquemas TVD

Uno de los objetivos de los métodos numéricos es lograr precisión en el tiempo y convergencia a soluciones físicamente correctas. Para este tipo de problemas es posible que surjan discontinuidades incluso cuando el problema es lineal y las condiciones iniciales son suaves. Además, algunos métodos pueden presentar oscilaciones en una región donde aparezcan gradientes pronunciados o discontinuidades. Así, es necesario asegurar comportamientos no oscilatorios o monótonos cerca de estas zonas para obtener resultados físicamente correctos. Por esta razón se construyen esquemas de alto orden que preserven la monotonicidad, de acuerdo a la siguiente definición (LeVeque, 1992; Mattheij et al., 2005).

Definición 4.1. Decimos que un esquema numérico para la ley de conservación (4.5) preserve la monotonicidad si se cumple lo siguiente: si el conjunto $\{u_j^n\}$ de soluciones en los nodos en el paso t^n es monótono creciente (decreciente), entonces el conjunto $\{u_j^{n+1}\}$ de soluciones al paso t^{n+1} también es monótono creciente (decreciente).

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = S(x, t, u). \quad (4.5)$$

Por otro lado, para medir el comportamiento oscilatorio en la solución numérica $\mathbf{u}^n = \{u_j^n\}$

definimos la variación total de ésta, $TV(\mathbf{u}^n)$, como

$$TV(\mathbf{u}^n) = \sum_{j=1}^{Ne} |u_{j+1}^n - u_j^n|. \quad (4.6)$$

Así, buscamos métodos que no aumenten las oscilaciones de un paso de tiempo a otro. Para esto, tenemos la siguiente definición (Shu, 1988; Shu y Osher, 1988, 1989).

Definición 4.2. Un esquema numérico para la ley de conservación (4.5) disminuye la variación total si

$$TV(\mathbf{u}^{n+1}) \leq TV(\mathbf{u}^n),$$

para cualquier solución \mathbf{u}^n y $n > 0$. En este caso decimos que el esquema es TVD.

Pedir que un método sea TVD es menos restrictivo que pedir que preserve la monotonicidad, pero al mismo tiempo, permite encontrar esquemas que la mantengan. Esto gracias al siguiente teorema.

Teorema 4.2.1. *Un esquema TVD para una ley de conservación preserva la monotonicidad.*

Una demostración del teorema puede consultarse en (LeVeque, 1992).

En (Cockburn y Shu, 2001) se demuestra que podemos obtener estabilidad de esquemas Runge-Kutta-TVD de alto orden a partir de la estabilidad del método de Euler hacia adelante. Para esto, el método de Runge-Kutta de orden $k + 1$ es reescrito como sigue:

1. Sea $u_h^0 = u(x, 0)$
2. Para $n = 0, 1, \dots, N - 1$ calcular u^{n+1} a partir de u^n de la siguiente manera:
 - a) $u_h^{(0)} = u_h^n$.
 - b) Para $i = 1, \dots, k + 1$ calcular las etapas intermedias:

$$u_h^{(i)} = \sum_{l=0}^{i-1} \left(\alpha_{il} u^{(l)h} + \Delta t^n \beta_{il} L(u_h^{(l)}) \right), \quad \alpha_{il} \geq 0, \quad \beta_{il} \geq 0, \quad i = 1, \dots, k + 1.$$

- c) $u^{n+1} = u^{(k+1)}$.

Y se presenta el teorema

Teorema 4.2.2. *Asumiendo que el método de Euler hacia adelante es TVD y estable bajo la restricción CFL $\Delta t < \delta_0$, con*

$$\delta_0 = \max_{0 \leq n \leq N} \{ \Delta t^n \max \{ \beta_{il} / \alpha_{il} \} \}$$

y que

$$\sum_{l=0}^{i-1} \alpha_{il} = 1, \quad i = 1 \dots k + 1,$$

entonces el método de Runge-Kutta de orden $k + 1$ (4.2) también es TVD y estable bajo la misma restricción CFL.

Observemos que el método de Runge-Kutta (4.4) se debe reescribir en la forma (4.2) para que satisfaga las hipótesis del teorema 4.2.2, implicando que sea TVD y estable. Esta propiedad es muy importante, ya que permite obtener estabilidad para métodos de alto orden a partir de la estabilidad del método de Euler hacia adelante. En particular, si la solución de los elementos de GD se aproximan por polinomios de grado k , entonces se debe de utilizar un RK-TVD de orden $k + 1$.

Cabe mencionar que en (Chavent y Cockburn, 1989) se muestra que para el caso lineal, $f(u) = cu$, usando una aproximación lineal, a trozos, es decir, $k = 1$, el método de Euler hacia adelante es inestable. Por otro lado, en (Cockburn y Shu, 1991) se muestra que si un Runge-Kutta de segundo orden se utiliza, el esquema es estable siempre que

$$c \frac{\Delta t}{\Delta x} \leq \frac{1}{3}.$$

Ver resultados numéricos en el apartado 5.1.

Por su parte, el esquema de Runge-Kutta-TVD de segundo orden (RK-TVD) (4.4) se puede escribir como

$$u_h^{(1)} = u_h + \Delta t L(u_h^n), \quad (4.7)$$

$$u_h^{(2)} = u_h^{(1)} + \Delta t L(u_h^{(1)}), \quad (4.8)$$

$$u_h^{n+1} = \frac{1}{2} (u_h^n + u_h^{(2)}). \quad (4.9)$$

La idea detrás de la construcción de esquemas TVD es que las soluciones físicas para problemas hiperbólicos no permiten la aparición de un nuevo extremo con el paso del tiempo. Es por esto que en los esquemas TVD la variación total de la solución numérica está controlada en alguna forma para prevenir la aparición de cualquier falso extremo. Así, el problema principal en el diseño de los esquemas TVD es introducir mecanismos para mantener la monotonicidad sin perder la precisión lejos de discontinuidades. Estos mecanismos pueden presentarse en forma de limitadores de pendiente o dentro de la función de flujo. En el siguiente apartado presentamos una de estas técnicas, los limitadores de pendiente.

Además, los esquemas TVD tienen problemas cuando los datos iniciales tienen extremos, llevando a disipación en la solución. Este problema puede superarse si pedimos que el esquema satisfaga la siguiente definición.

Definición 4.3. Un método es de variación total acotada (es TVB) si para cualquier conjunto

de datos iniciales $\mathbf{u}^0 = \{u_j^0\}$, con $TV(\mathbf{u}^0) < \infty$, y cualquier tiempo T , existe una constante $M > 0$ y un valor $\delta > 0$ tal que

$$TV(u^n) \leq M,$$

para todo $n\Delta t \leq T$ siempre que $\Delta t < \delta$.

Claramente todo esquema TVD es automáticamente TVB, tomando $M = TV(\mathbf{u}^0)$.

4.3. Limitadores de pendiente

A pesar que los esquemas sean TVD o TVBM teóricamente, esto no garantiza que la solución numérica se comporte como se espera, ya que oscilaciones espurias pueden aparecer cerca de discontinuidades o de pendientes pronunciadas, por ejemplo en la ecuación de convección con declives muy pronunciados. Para evitar que esto suceda, se emplean métodos que aseguran la estabilidad de la variación total directamente. Estos métodos se conocen como limitadores de pendiente. Cabe mencionar que estos limitadores mantienen la propiedad de ser TVD mientras las condición CFL se satisfaga.

Supongamos que la discretización del tiempo se realiza por medio de Euler hacia adelante (4.2). Sea $u_e^{n,pre}$ una solución preliminar en el elemento Ω_e obtenida con el método de Galerkin discontinuo al tiempo t^n , donde $u_e^{0,pre}$ es la condición inicial sobre ese elemento. La idea de un limitador de pendiente es reconstruir ésta solución preliminar usando los elementos vecinos, para obtener una solución, $u_e^{n,lim}$, en la que las oscilaciones se hayan eliminado antes de avanzar al siguiente paso de tiempo.

El proceso de limitación debe satisfacer las siguientes condiciones:

1. No debe cambiar el promedio de la celda de $u_e^{n,pre}$. Es decir, los promedios en las celdas de $u_e^{n,pre}$ y $u_e^{n,lim}$ deben ser los mismos. Esto para mantener la propiedad de conservación del método de Galerkin discontinuo.
2. No debe afectar la precisión del esquema en regiones suaves.
3. Debe eliminar las oscilaciones cerca de zonas con pendientes pronunciadas o de discontinuidades.

El promedio de celda de la solución $u_e^{n,pre}$ está definido por

$$\bar{u}_e = \frac{1}{\Delta x} \int_{\Omega_e} u_e^{n,pre} dx, \quad (4.10)$$

La solución limitada en el elemento está representada por la expresión

$$u_e^{n,lim} = \bar{u}_e + \sigma_e(x - \bar{x}), \quad x \in [x_l^e, x_r^e]$$

donde \bar{x} es el punto medio del elemento $[x_l^e, x_r^e]$ y σ_e es la pendiente limitada sobre el elemento.

Para calcular σ_e utilizamos la función *minmod* dada por

$$\sigma_e = m(a, b, c) = \begin{cases} s \min(|a|, |b|, |c|) & \text{si } s = \text{sign}(a) = \text{sign}(b) = \text{sign}(c), \\ 0 & \text{otro caso,} \end{cases} \quad (4.11)$$

donde los argumentos de la función, a , b y c , son la pendiente hacia atrás (en Ω_{e-1}), la pendiente hacia delante (en Ω_{e+1}) y la pendiente central (en Ω_e), respectivamente, definidas por

$$a = \frac{\bar{u}_e - \bar{u}_{e-1}}{\bar{x}_e - \bar{x}_{e-1}}, \quad b = \frac{\bar{u}_{e+1} - \bar{u}_e}{\bar{x}_{e+1} - \bar{x}_e}, \quad c = \frac{\bar{u}_{e+1} - \bar{u}_{e-1}}{\bar{x}_{e+1} - \bar{x}_{e-1}}. \quad (4.12)$$

La función *minmod* funciona de la siguiente manera: si las pendientes dadas en (4.12) tienen el mismo signo, regresa el valor de aquella que tenga el menor modulo, en otro caso, el valor que regresa es cero. Es esta función la que proporciona el mecanismo para decidir si la solución es suave o existen oscilaciones en un elemento al comparar la pendiente de la solución con las pendientes en los vecinos. Un esquema de esto puede verse en las figuras 4.1 y 4.2, donde las líneas continuas negras en las figuras a) y c) son las soluciones preliminares del método, $u_h^{n,pre}$, las líneas continuas negras en las gráficas b) y d) son las soluciones limitadas, $u_h^{n,lim}$, el promedio del celda de la solución, \bar{u}_e , se representa con líneas punteadas rojas.

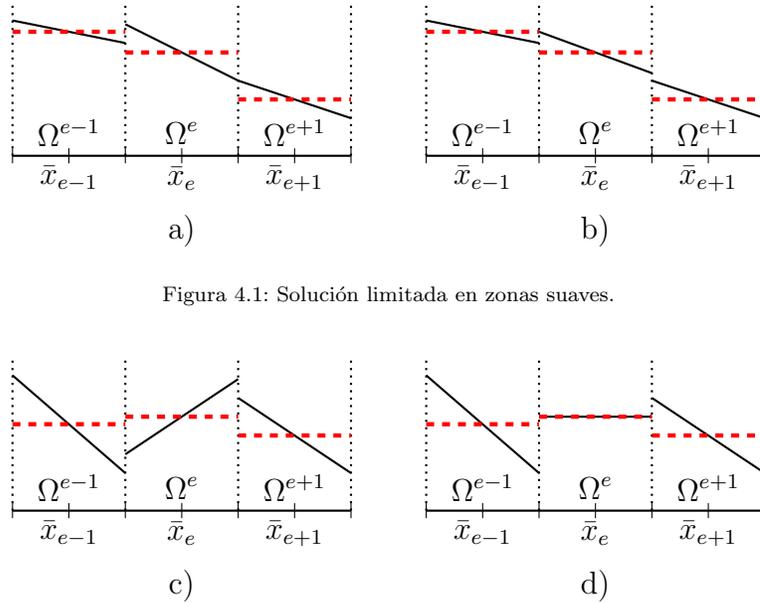


Figura 4.1: Solución limitada en zonas suaves.

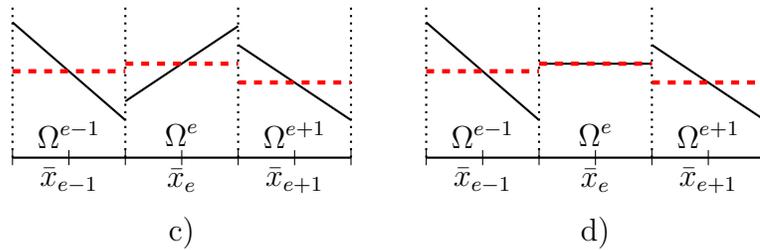


Figura 4.2: Solución limitada en zonas con oscilaciones

Para mantener la propiedad TVD del método de Euler hacia adelante definimos la variación total en las medias como

$$TVM(u_h) = \sum_{j=1}^{Ne} |\bar{u}_{j+1}^n - \bar{u}_j^n|, \quad (4.13)$$

y se tiene el siguiente resultado

Proposición 4.3.1. *Para condiciones de frontera con soporte compacto o periódicas, la solución numérica u_h^n del esquema Galerkin discontinuo con Euler hacia adelante (4.2) y limitada por la función $minmod$, es un esquema que disminuye la variación total en las medias (TVDM). Es decir*

$$TVM(u_h^{n+1}) \leq TVM(u_h^n)$$

Este resultado, aunque se presenta para esquemas en los que la discretización del tiempo se hace con Euler hacia adelante, también puede aplicarse a esquemas en los que se discretiza el tiempo con algún método RK-TVD, permitiendo obtener el mismo resultado de estabilidad cuando se utiliza el limitador $minmod$.

Proposición 4.3.2. *Bajo las mismas condiciones que en la proposición 4.3.1, la solución u_h^n del esquema Galerkin discontinuo con algún RK-TVD y limitada es TVDM.*

Sin embargo, el limitador elimina precisión cerca de extremos de la solución. Esto debido a que, si un elemento contiene un punto extremo, las pendientes hacia adelante y hacia atrás tendrán signo contrario, provocando que en esta región la función $minmod$ regrese cero, metiendo difusión en la solución. Para evitar esto, se utiliza un corrector que mantenga la variación total acotada en las medias (TVBM).

$$\tilde{m}(a_1, \dots, a_p) = \begin{cases} a_1 & \text{si } |a_1| \leq Mh^2, \\ m(a_1, \dots, a_p) & \text{otro caso,} \end{cases}, \quad (4.14)$$

donde el parámetro M es una cota superior del valor absoluto de la segunda derivada evaluada en los extremos de la solución (Cockburn y Shu, 2001; Hesthaven y Warburton, 2007). En el caso de leyes de conservación no lineales y si la condición inicial, $u_0(x)$, es \mathcal{C}^2 a trozos, entonces

$$M = \sup \{|u_0''(y)|, : u_0'(y) = 0\}. \quad (4.15)$$

Con esto, el esquema será de variación total acotada en las medias (TVBM) y de alto orden de precisión para soluciones con extremos. Más detalles en (Cockburn y Shu, 2001). A esta corrección, ecuación (4.14) se le llama limitador de pendiente TVBM.

Finalmente, en (Cockburn y Shu, 2001) se sugiere aplicar los limitadores en cada etapa intermedia del método de RK-TVD, dado que funcionan mejor que aplicarlos a la solución de la última etapa.

Capítulo 5

Resultados Numéricos

En este capítulo presentamos los resultados obtenidos al aplicar el método de Galerkin discontinuo (GD). Primero, probamos la eficacia del método GD con Euler-TVD y Runge-Kutta-TVD de segundo orden y los limitadores, tanto en precisión como en la captura de ondas de choque y de rarefacción. Para esto resolvemos la ecuación de convección y el problema de Riemann para la ecuación de Burgers. Posteriormente, utilizamos estas técnicas junto con los flujos numéricos HLL, Roe y Lax para encontrar la solución de las ecuaciones de aguas someras en diversos ejemplos como: fractura de presa sobre una cama plana, mojada o seca, con o sin fricción; salto hidráulico; tránsito de agua sobre un tope con diferentes regímenes o estados de flujo; movimiento de agua sobre un lecho irregular y sobre un bol parabólico.

Los primeros ejemplos fueron resueltos en una Laptop HP con sistema operativo Windows 8, con un procesador Intel Celeron N2840 a 2.16GHz y con 2 GB de RAM, en la versión R2015a de MATLAB. Los problemas de flujo sobre un tope fueron ejecutados en una computadora de escritorio Dell, con procesador Intel Core i5-4590 a 3.30GHz y con una memoria RAM de 4GB. Debido a que fue necesario más velocidad de cálculo, los ejemplos de cama irregular y de flujo sobre un bol se ejecutaron en una Dell con procesador Intel Core i7-6700 a 3.40GHz con una memoria RAM de 8GB.

5.1. Análisis de limitadores y flujos numéricos

5.1.1. Ecuación de convección

Con el fin de probar la efectividad de los limitadores de pendiente aplicamos el método de Galerkin discontinuo a la ecuación de convección con velocidad constante u :

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0, \quad x \in [0, 1], \quad t > 0,$$

con condición inicial

$$C(x, 0) = \begin{cases} \sin(10\pi x) & \text{si } x \in [0, 0.1] \\ 0 & \text{otro caso} \end{cases},$$

donde el número de elementos, ne , es igual a 100, 500 y 1000. Tomamos como criterio de parada que la solución supere una tolerancia para el error relativo igual a $TOL = 0.4$ para Euler hacia adelante y para el resto de los esquemas cuando se supere el tiempo final de simulación igual a $t = 5s$. En las figuras 5.1 a la 5.5 se muestran a la izquierda las gráficas de las soluciones analítica y numérica con 100, 500 y 1000 elementos de arriba a abajo, en el centro el máximo error relativo a lo largo del tiempo y a la derecha las gráficas de la variación total para cada caso, ver ecuación (4.6). El número de Courant-Friedrichs-Lewy utilizado para todos los casos es $CFL = 0.3$. Además, para el flujo numérico en las fronteras se ha utilizado un esquema *upwind*:

$$\tilde{c}^e = \begin{cases} c_l^e, & u > 0, \\ c_r^e, & u < 0. \end{cases}$$

Cuando utilizamos el método de Euler hacia adelante para integrar el tiempo sin un limitador obtenemos soluciones con oscilaciones, estas se generan en las zonas donde cambia la pendiente drásticamente, fig. 5.1. Las oscilaciones continúan aunque se tome un número mayor de elementos. Más aún, podemos ver que el error supera la tolerancia más temprano conforme se toman más elementos. Claramente la solución no cumple ser TVD.

Al utilizar el limitador de pendiente *minmod*, ecs. (4.11) y (4.12), en el método de Euler la solución mejora, pero observamos que supera la tolerancia al usar pocos elementos. Para lograr una mejor solución aumentamos el número de ellos, pero las soluciones presentan aglomeraciones en algunas zonas en vez de distribuirse uniformemente, fig. 5.2. La solución obtenida es TVD, como se espera, reafirmando que el limitador elimina las oscilaciones cercanas a zonas abruptas.

La figura 5.3, muestra la solución obtenida cuando aplicamos el método de Runge-Kutta-TVD de segundo orden. Esta solución es mejor que las anteriores pero nuevamente presenta ligeras oscilaciones en las zonas donde cambia la pendiente, específicamente en el frente de la solución. Conforme aumenta el número de elementos éstas logran desaparecer, aunque no del todo, provocando un aumento del tiempo de ejecución y la cantidad de memoria utilizada. Nuevamente, el método sin el limitador no es TVD.

Con el fin de quitar las oscilaciones aplicamos el limitador *minmod*, como se puede observar en la figura 5.4, pero aparece algo de difusión numérica provocando que el error aumente. Esta difusión es provocada por la existencia de un máximo en la condición inicial. Un aumento en el número de los elementos permite disminuir la difusión numérica.

Ahora, para evitar la difusión numérica usamos el limitador de pendiente TVBM. Con ello logramos que la difusión desaparezca cuando se trabaje con pocos elementos, figura 5.5,

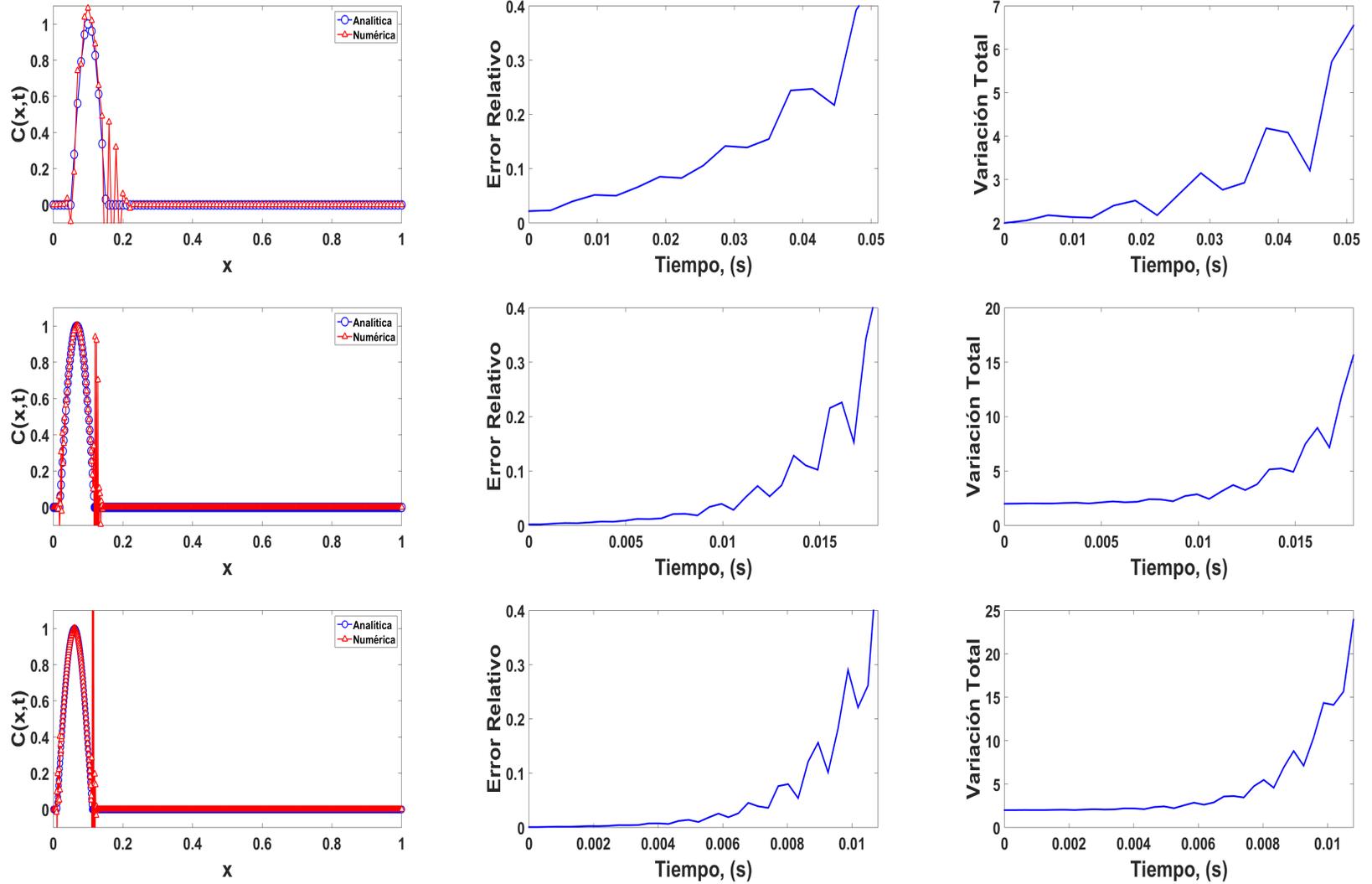


Figura 5.1: Solución de la ecuación de convección usando el método de Euler hacia adelante sin limitador de pendiente después de que el error superó una $TOL = 0.4$ (derecha), máximo error relativo por paso de tiempo (centro) y variación total (izquierda) en cada paso de tiempo con número de elementos igual a 100 (arriba), 500 (en medio) y 1000 (abajo).

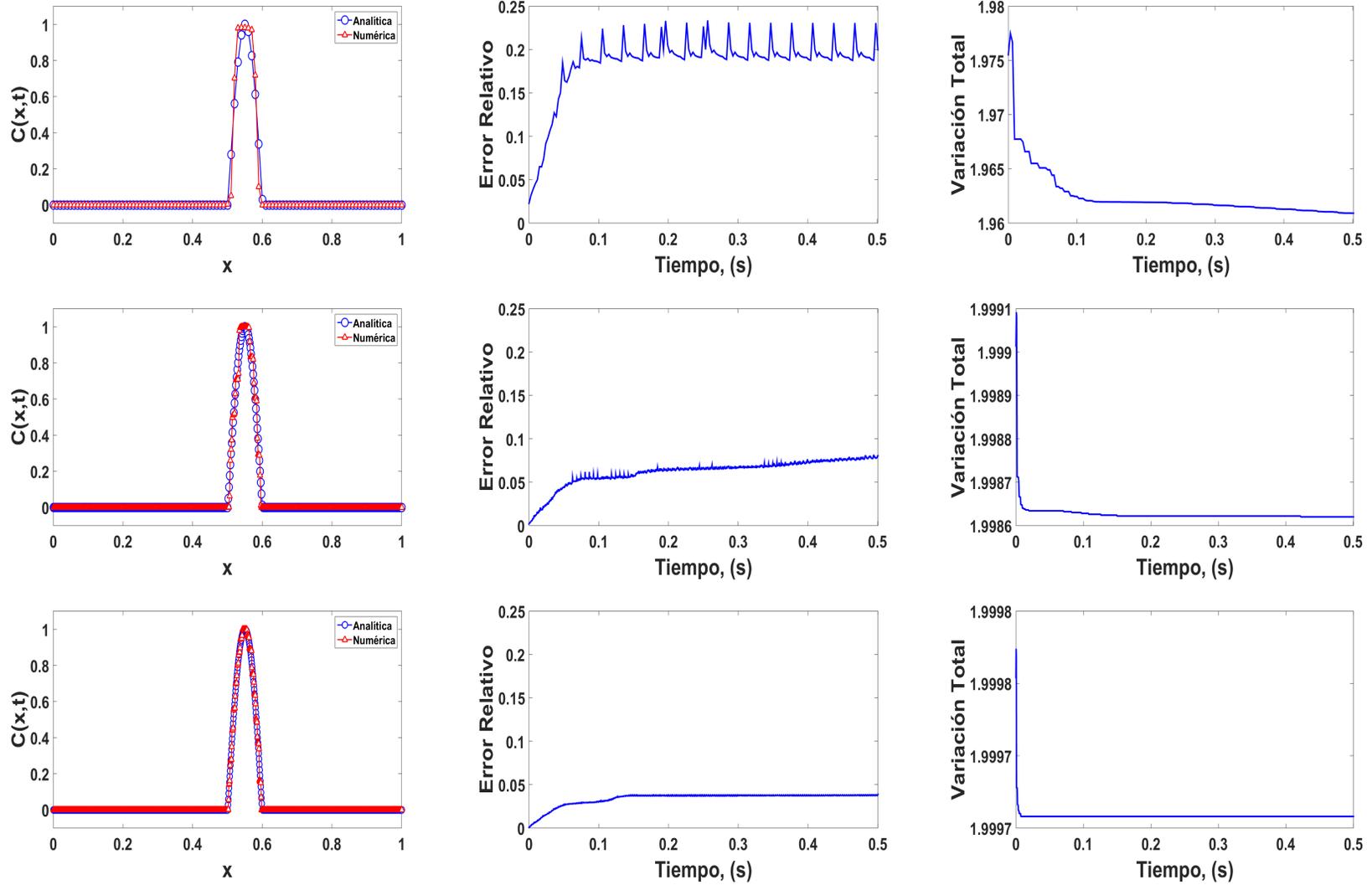


Figura 5.2: Solución de la ecuación de convección después de 0.5 segundos (derecha), error (centro) y variación total (izquierda) usando Euler hacia adelante y el limitador minmod con número de elementos 100 (arriba), 500 (en medio) y 1000 (abajo).

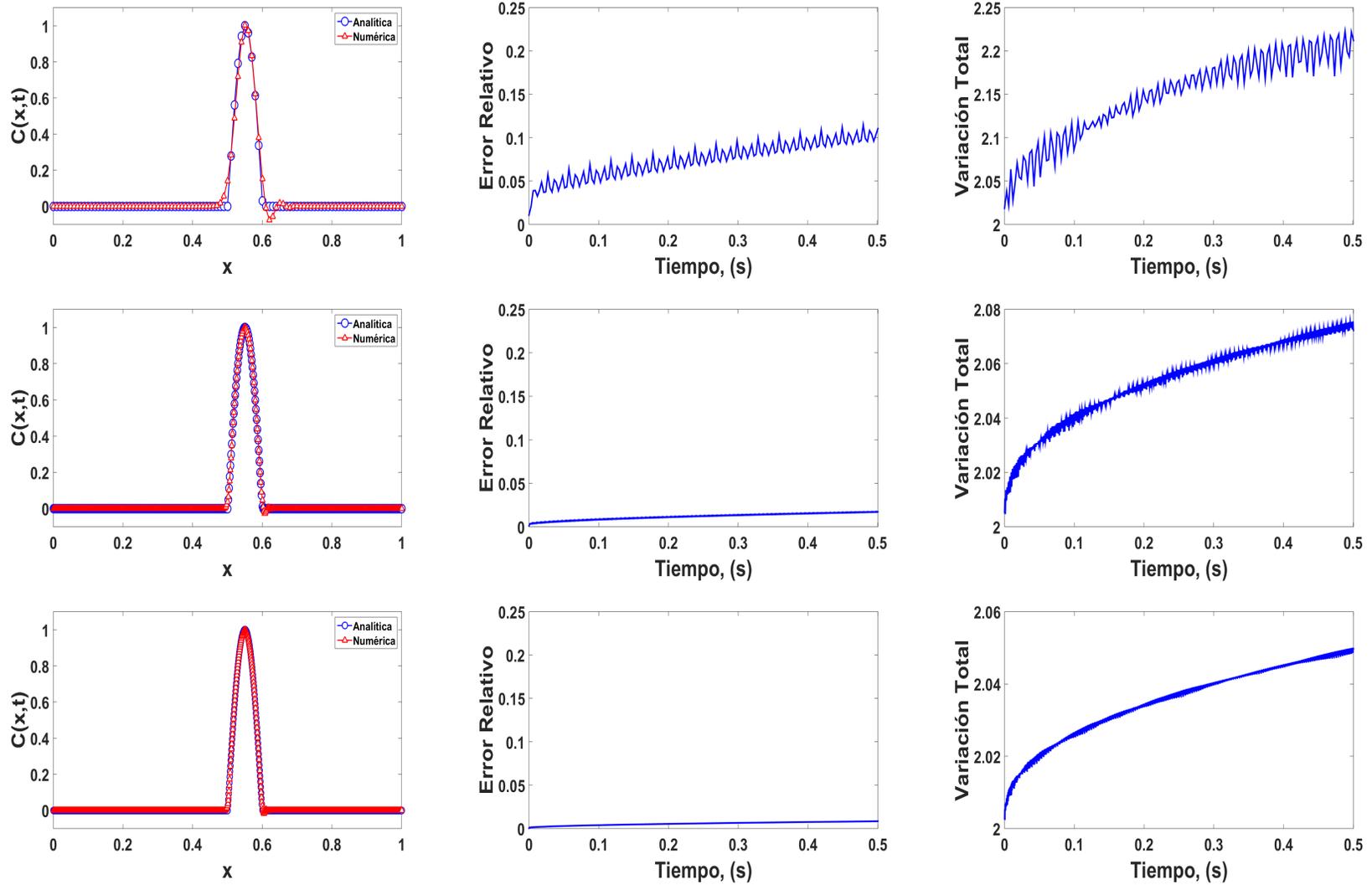


Figura 5.3: Solución de la ecuación de convección después de 0.5 segundos (derecha), error (centro) y variación total (izquierda) usando RK-TVD de segundo orden con número de elementos 100 (arriba), 500 (en medio), 1000 (abajo).

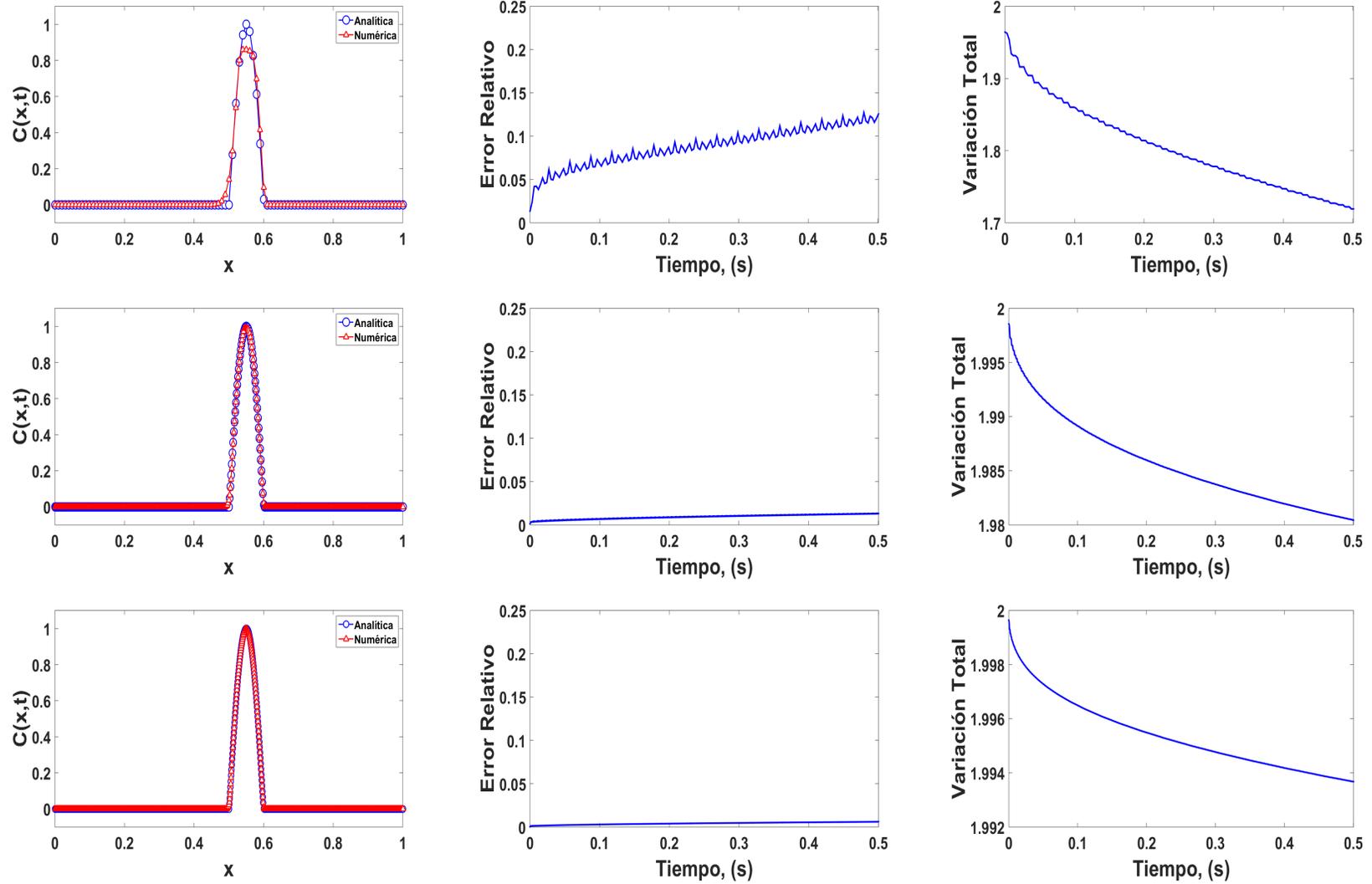


Figura 5.4: Solución de la ecuación de convección después de 0.5 segundos (derecha), error (centro) y variación total (izquierda) usando RK-TVD de segundo orden y el limitador minmod con número de elementos 100 (arriba), 500 (en medio), 1000 (abajo).

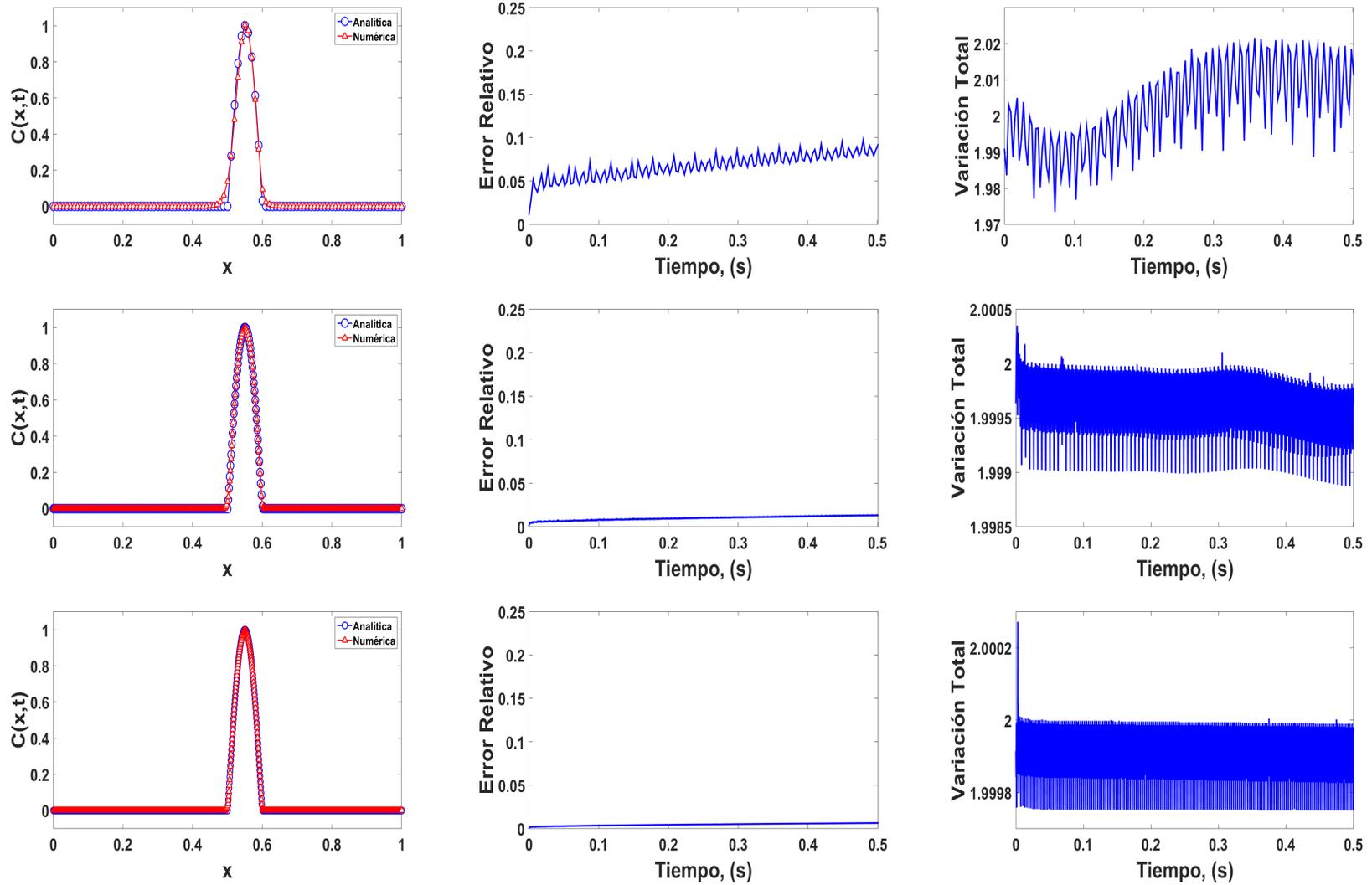


Figura 5.5: Solución de la ecuación de convección después de 0.5 segundos (derecha), error (centro) y variación total (izquierda) usando RK-TVD de segundo orden y el limitador de pendiente TVBM. Número de elementos 100, 500, 1000.

disminuyendo así el tiempo de ejecución. Al observar la variación total, podemos notar que aunque esté oscilando con el tiempo se mantiene acotada. Dado que la solución es continua en cada elemento, alcanza su máximo y su mínimo, por lo cual el valor promedio va a estar acotado por estos valores, implicando que la variación total en las medias esté acotada, esto es, la solución satisface ser TVBM.

Al observar las gráficas de errores relativos para los métodos de Euler y RK-TVD con 500 elementos, ver 5.6, podemos concluir que el método de RK-TVD sin limitador es mejor que el método de Euler hacia adelante con el limitador minmod. Además, cuando se aplican los dos tipos de limitadores al método de RK-TVD, los resultados son mejores que Euler limitado, teniendo el mismo comportamiento, ver figura 5.7.

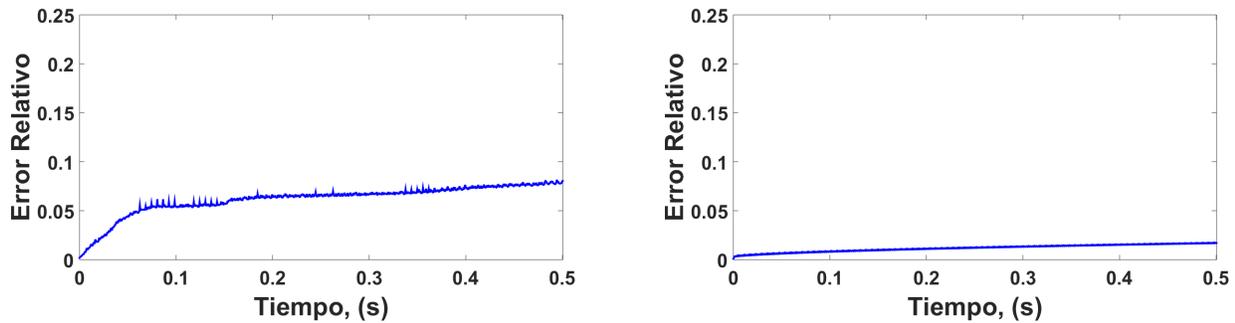


Figura 5.6: Máximo de los errores relativos contra el tiempo de simulación para los métodos de Euler-TVD con limitador minmod (izquierda) y RK-TVD de segundo orden sin limitador (derecha).

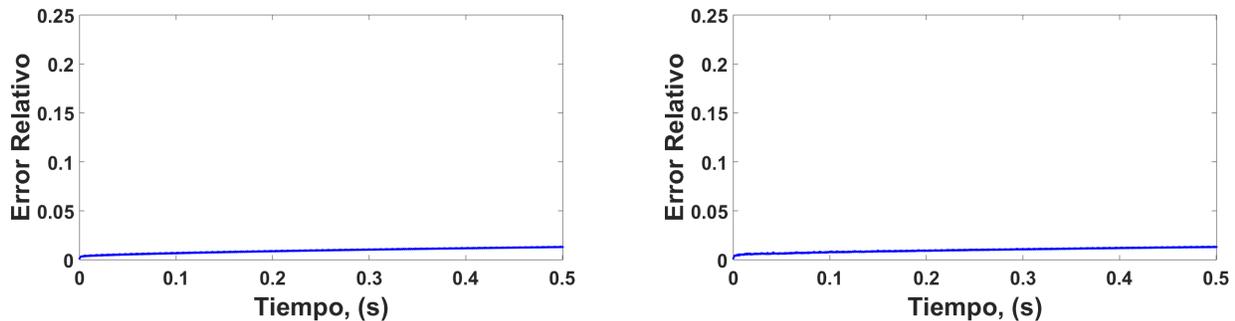


Figura 5.7: Máximo de los errores relativos contra el tiempo de simulación para el método de RK-TVD con el limitador minmod (izquierda) y el limitador TVBM (derecha).

El máximo número de CFL es $1/3$. Si se supera, empiezan a aparecer oscilaciones que provocan que la solución sea inestable a pesar de aplicar algún limitador de pendiente.

Por último, la tabla 5.1 muestra los tiempos de ejecución de los dos métodos sin y con limitador para CFL iguales a 0.1 y 0.3 en minutos. Notemos que cuando se ocupan 500 elementos el tiempo disminuye a la mitad al aumentar el CFL, mientras que con 1000 elementos

la diferencia en tiempo es pequeña e incluso en algunos casos llega a ser mayor. Como conclusión final, el mejor resultado fue el obtenido con el método de RK-TVD con cualquier limitador y 500 elementos, ya que este ocupa poco tiempo de ejecución y memoria, además de obtener una solución con una precisión similar a la dada con mallas más finas.

Tiempos de ejecución para los diferentes métodos

CFL	Euler	Euler <i>minmod</i>	RK-TVD	RK-TVD <i>minmod</i>	RK TVBM	No. Pasos de tiempo	dt
100 elementos							
0.1	1.65	1.54	2.43	1.61	1.71	200	0.001
0.3	0.60	0.54	0.81	0.54	0.55	66.67	0.003
500 elementos							
0.1	8.96	8.80	9.96	10.14	10.36	1000	0.002
0.3	3.07	3.24	4.21	3.89	3.79	333.33	0.006
1000 elementos							
0.1	8.44	7.98	10.20	10.29	11.47	666.67	0.0003
0.3	7.38	8.38	9.35	9.75	12.37	600	0.0003333

Tabla 5.1: Tiempos de ejecución en minutos para $n_e = 100, 500$ y 1000 .

5.1.2. Ecuación de Burgers

Como segundo ejemplo, resolvemos la ecuación de Burgers

$$\frac{\partial u}{\partial t} + \frac{1}{2} \frac{\partial u^2}{\partial x} = 0, \quad x \in [-1, 1], \quad t > 0,$$

con condición inicial discontinua a trozos

$$u(x, 0) = u_0(x) = \begin{cases} u_l & x < 0 \\ u_r & x \geq 0 \end{cases}$$

para observar la eficacia del método en la captura de ondas de choque ($u_l > u_r$) y rarefacción ($u_l < u_r$). Estos fenómenos son importantes pues también se presentan en el modelo de las ecuaciones de aguas someras.

El dominio se ha discretizado con 300 elementos y se muestran los resultados para $t = 0.6s$ en el caso de onda de rarefacción y de $t = 3s$ para onda de choque.

Además, puesto que se trata de una ecuación no lineal, el número CFL que se utiliza tiene la forma siguiente:

$$CFL = \frac{\Delta t \max(\lambda)}{\Delta x}$$

donde λ es el valor propio de la ecuación evaluado en las condiciones iniciales, ver (Mattheij

et al., 2005). Así, la condición CFL se mantuvo igual a 0.2 para onda de choque y a 0.1 para onda de rarefacción.

En las figuras 5.8 a la 5.10 comparamos los resultados entre el Runge-Kutta tradicional (gráficas de la izquierda) y el RK-TVD (gráficas de la derecha) ambos de segundo orden. En los ejemplos de convección, figura 5.8 y en el de la ecuación de Burgers con onda de rarefacción, figura 5.9, no se presenta algún cambio. Sin embargo, para la onda de choque de la ecuación de Burgers observamos que aparecen oscilaciones en la solución con el RK no TVD, figura 5.10. Esto muestra la importancia de los métodos TVD en aquellos problemas con soluciones discontinuas.

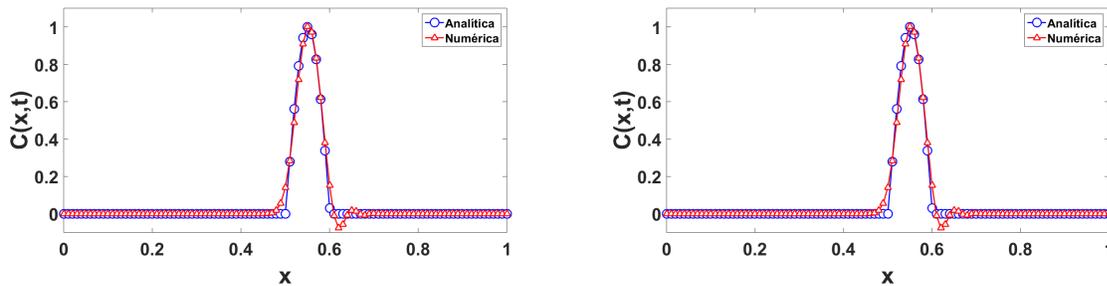


Figura 5.8: Solución de la ecuación de convección con los métodos de RK tradicional (izquierda) y RK-TVD (derecha)

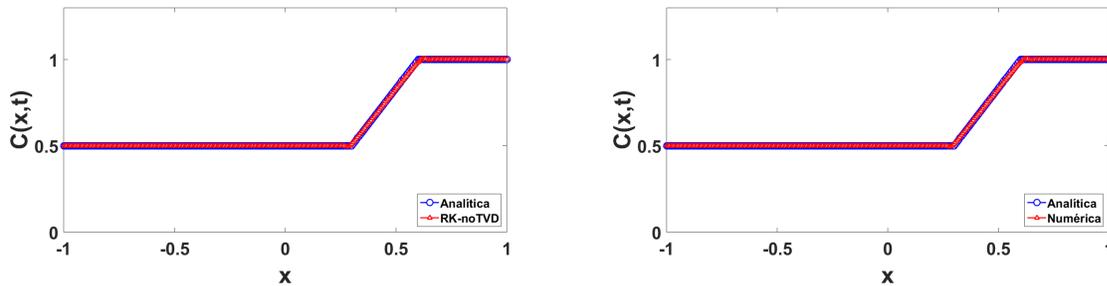


Figura 5.9: Solución de la ecuación de Burgers con onda de rarefacción utilizando los métodos de RK tradicional (izquierda) y RK-TVD (derecha)

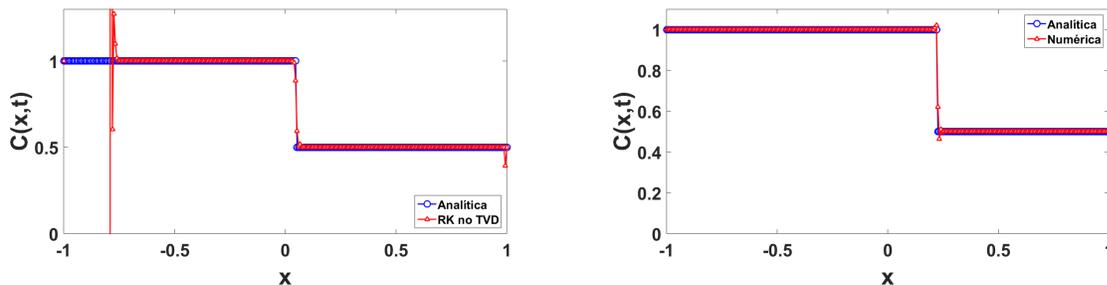


Figura 5.10: Solución de la ecuación de Burgers con onda de choque utilizando los métodos de RK tradicional (izquierda) y RK-TVD (derecha)

En las figuras 5.11 a la 5.14 se presentan las gráficas de concentración contra el espacio (izquierda), error relativo contra tiempo (centro) y variación total contra tiempo (derecha); además de arriba hacia abajo los resultados del método sin limitador, en medio el método con el limitador *minmod* y abajo con el limitador TVBM.

En la figura 5.11 presentamos los resultados para la solución de onda de choque, caso $u_l > u_r$, con el método de Euler hacia adelante TVD. Notamos que en general el método de Euler captura bien la discontinuidad que se propaga a lo largo de la característica. Al aplicar los limitadores *minmod* y TVBM, la solución no cambia significativamente. La variación total en estos últimos permanece constante. Eso verifica que no aparecen falsas oscilaciones. No se aprecia una gran diferencia entre ambos limitadores.

La figura 5.12 contiene los resultados con el método de Runge-Kutta-TVD. Claramente al usar un número de CFL más pequeño mejora la solución. Aunque llegan a aparecer pequeñas oscilaciones en la discontinuidad. Al quitar estas con alguno de los limitadores se pierde algo de precisión en la solución.

El número CFL máximo para que el método se mantuviera estable en ondas de choque es de 0.5 para RK-TVD.

Para el caso de onda de rarefacción, $u_l < u_r$, el método de Euler presenta oscilaciones que van aumentando conforme avanza el tiempo, figura 5.13, provocando que el máximo del error relativo incremente. Por otro lado, cuando se limita la solución, el comportamiento oscilatorio desaparece pero se presenta un ligero desfase en el frente de la solución. La solución mejora con el método de Runge-Kutta-TVD. Nuevamente, no se aprecian diferencias entre las soluciones limitadas, fig. 5.14. Los errores van disminuyendo conforme pasa el tiempo, esto debido a que la discontinuidad en la condición inicial va transformándose en una solución continua, conocida como solución de similaridad.

El máximo número CFL para mantener estabilidad con ondas de rarefacción es de 1/3 para RK-TVD.

A partir de los errores relativos podemos apreciar que la solución aproximada de cualquier método presenta características similares. De esto podemos inferir que la elección del método ocupado depende de si queremos mayor precisión (usando el método de RK-TVD) o un menor tiempo de ejecución (usando Euler-TVD). El tiempo máximo de ejecución para ambos casos es de 2 minutos.

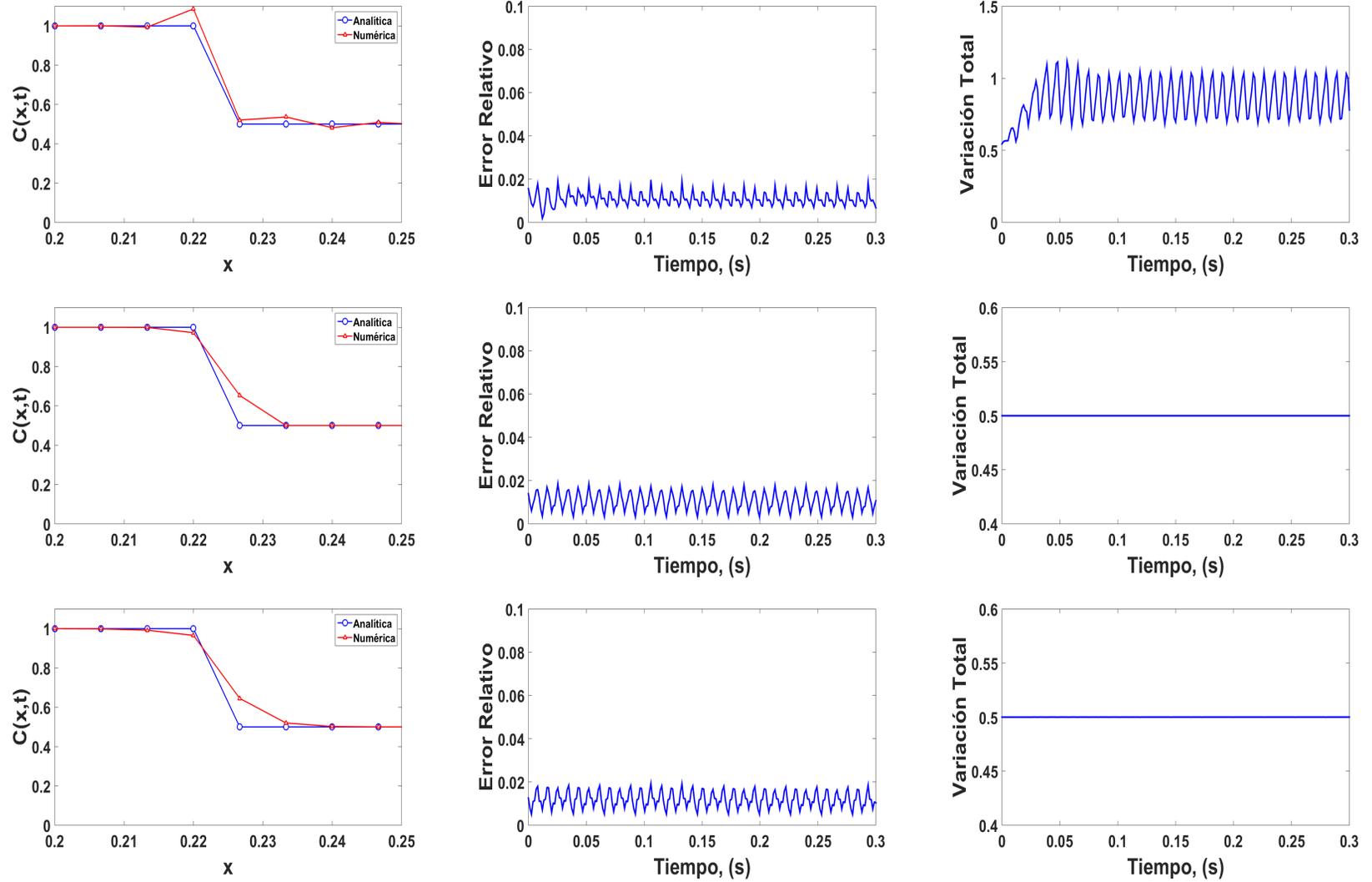


Figura 5.11: Solución con onda de choque para la ecuación de Burgers (derecha), errores relativos (centro) y variación total (izquierda) usando Euler-TVD sin limitador (arriba), con limitador $minmod$ (segunda fila) y con limitador TVBM (abajo) con 300 elementos y CFL=0.2.

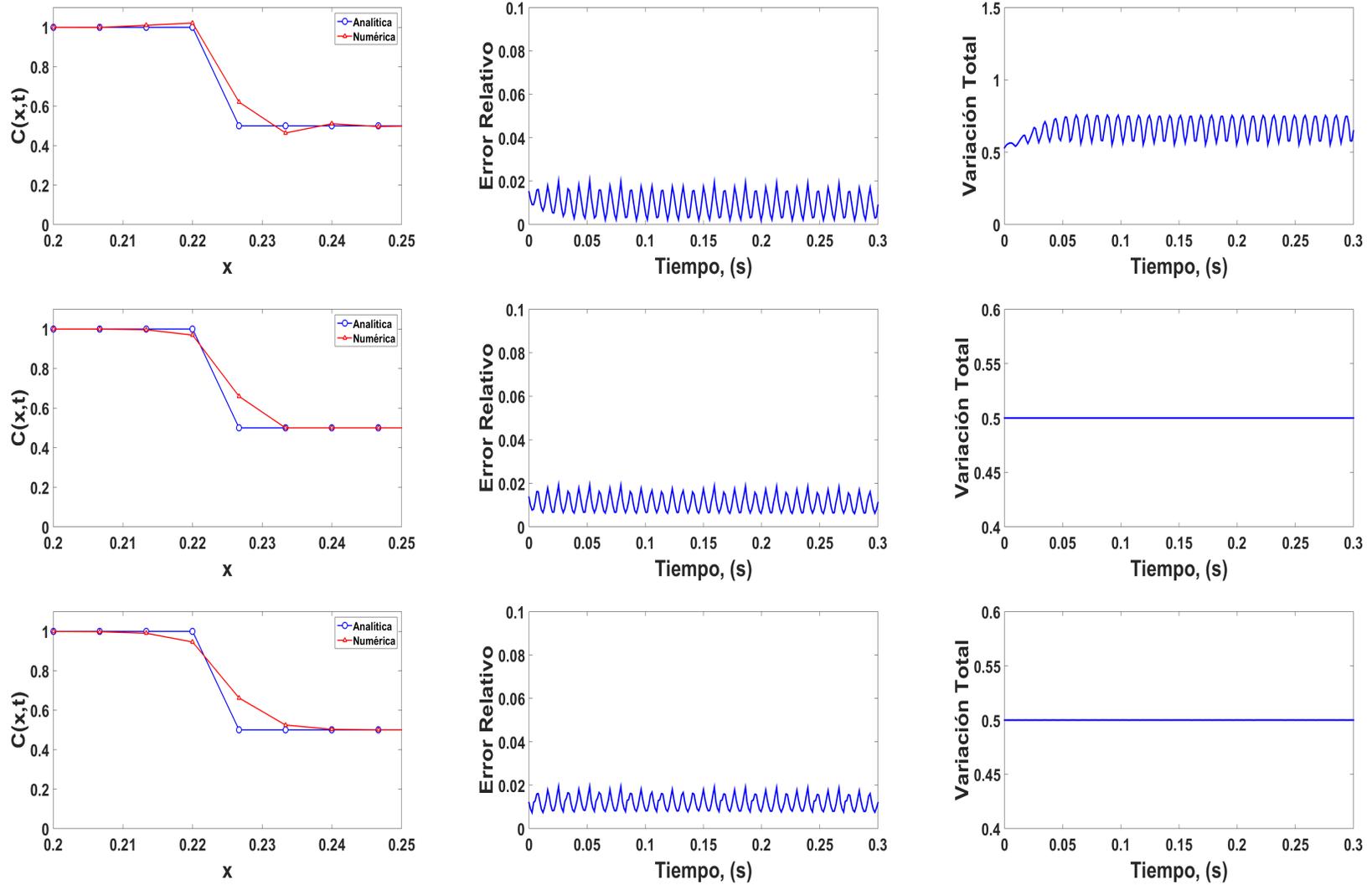


Figura 5.12: Solución con onda de choque para la ecuación de Burgers (derecha), errores relativos (centro) y variación total (izquierda) usando RR-TVD sin limitador (arriba), con limitador $minmod$ (segunda fila) y con limitador TVBM (abajo) con 300 elementos y CFL=0.2.

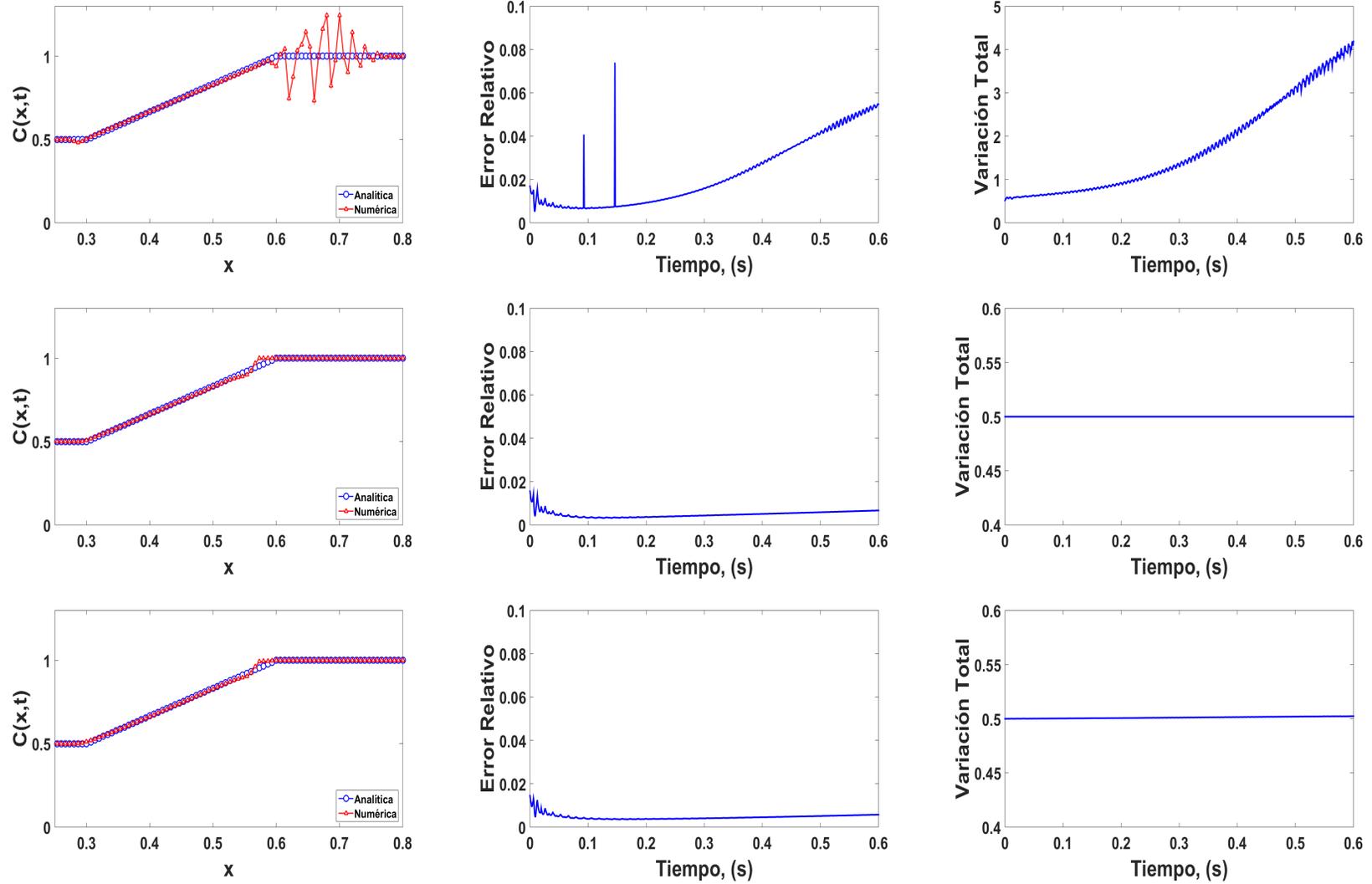


Figura 5.13: Solución para onda de rarefacción de la ecuación de Burgers (derecha), errores relativos (centro) y variación total (izquierda) usando Euler-TVD sin limitador (arriba), con limitador *minmod* (segunda fila) y con limitador TVBM (abajo) con 300 elementos y CFL=0.1

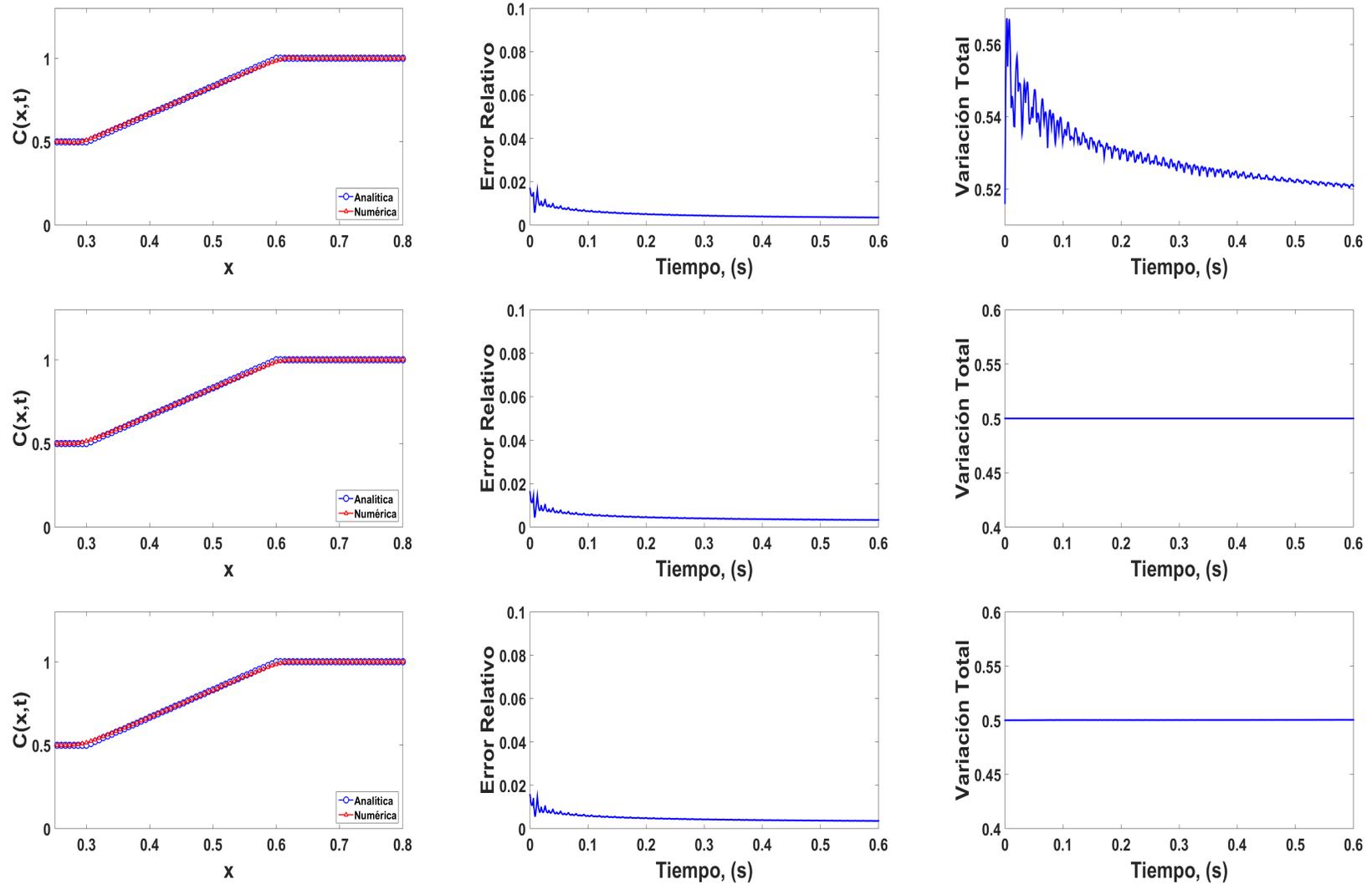


Figura 5.14: Solución para onda de rarefacción de la ecuación de Burgers (derecha), errores relativos (centro) y variación total (izquierda) usando RK-TVD sin limitador (arriba), con limitador $minmod$ (segunda fila) y con limitador TVBM (abajo) con 300 elementos y CFL=0.1

5.2. Ecuaciones de aguas someras

La metodología presentada en los apartados anteriores permiten decidir un procedimiento para encontrar la solución de las ecuaciones de aguas someras, esto tomando en consideración la precisión de las soluciones, la eliminación de oscilaciones y el tiempo de ejecución. Así se decidió usar los métodos de Euler hacia adelante y el método de Runge-Kutta-TVD de segundo orden ambos con el limitador *minmod*.

Primero se prueba con Euler hacia adelante, pues tarda menos en ejecutarse, y luego se utiliza Runge-Kutta-TVD de segundo orden para aumentar la precisión.

A continuación se presentan los resultados para las ecuaciones de aguas poco profundas. Al igual que con la ecuación de Burgers, el número de Courant que se utiliza depende de los valores propios y tiene la forma

$$C = \frac{\Delta t \max(\lambda_i)}{\Delta x}$$

donde λ_i son los valores propios de las ecuaciones de aguas someras, ecuaciones (2.23), evaluados en las condiciones iniciales.

5.2.1. Fractura de presa idealizada en un canal rectangular sobre lecho mojado

Se resuelve el problema de fractura de presa en un canal de 1000 m de longitud, con una presa localizada a 500 m. La profundidad del agua al inicio de la simulación dentro de la presa es de 10 m y fuera de ella es de 2 m, obteniendo así un problema de Riemann para las ecuaciones de aguas poco profundas. No consideramos la fricción en el canal y la pendiente del lecho no cambia, implicando una ecuación homogénea. La presa se fractura y el flujo de agua subsecuente es simulado. La solución analítica a este problema se detalló en el capítulo 2. Se utilizan 400 elementos para discretizar el dominio, un CFL igual a 1/3. La figura 5.15 presenta la solución después de 20 segundos con el método de Runge-Kutta-TVD de segundo orden y los flujos numéricos HLL, HLL-FT, LF y Roe. Los resultados numéricos logrados con los flujos HLL y Roe son los más cercanos a la solución exacta en las zonas donde aparecen las ondas de rarefacción, ver figura 5.16. Con los flujos HLL, HLL-FT y Roe, el número CFL puede crecer hasta 1/2 sin la aparición de irregularidades en la solución; por su parte, el de Lax-Friedrichs no permite que supere 1/3. Así, al utilizar los primeros para la simulación podemos dar un paso más grande de tiempo sin perder precisión.

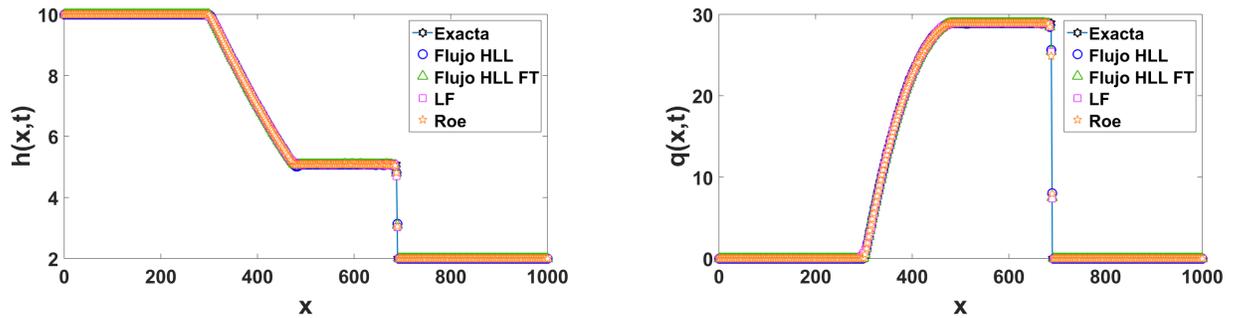


Figura 5.15: Solución para fractura de presa. Profundidad $h(x, t)$ (izquierda), caudal $q(x, t)$ (derecha).

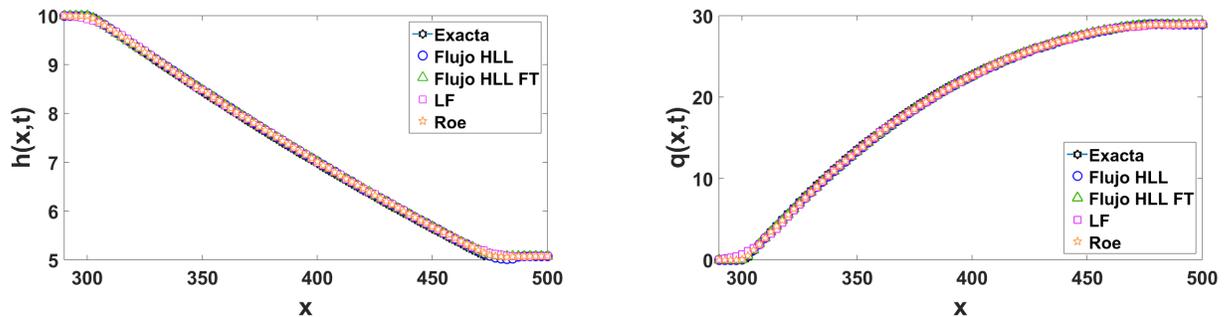


Figura 5.16: Onda de rarefacción para la solución de fractura de presa. Profundidad $h(x, t)$ (izquierda), caudal $q(x, t)$ (derecha).

5.2.2. Fractura de presa idealizada en un canal rectangular sobre lecho seco

El siguiente escenario para nuestras ecuaciones es análogo al anterior, es decir, al inicio la profundidad del agua dentro de la presa es de 10 m pero considerando ahora que fuera de la ella no hay agua. Este es el conocido problema mojado-seco.

Sabemos del capítulo 2, que las ecuaciones de aguas someras son hiperbólicas cuando $h > 0$. Un tratamiento para el problema de cama seca, $h = 0$, es necesario, ya que los flujos numéricos con solver de Riemann están basados en leyes de conservación hiperbólicas. Así, el flujo numérico con cama seca debe ser calculado adecuadamente y la profundidad del agua debe ser físicamente correcta, es decir, $h \geq 0$.

Dado un nodo y considerando que se encuentre en la frontera de la cama seca, las velocidades de onda, S_l y S_r , para el flujo HLL están determinadas por las siguientes ecuaciones (5.1), (Toro, 1990).

Si el nodo está seco a la derecha, es decir, $h_l > 0$ y $h_r = 0$, entonces

$$S_l = u_l - \sqrt{gh_l}, \quad S_r = u_l + 2\sqrt{gh_l}. \quad (5.1a)$$

En caso contrario, $h_l = 0$ y $h_r > 0$, entonces

$$S_l = u_r - 2\sqrt{gh_r}, \quad S_r = u_l + \sqrt{gh_r}. \quad (5.1b)$$

De igual manera, si ocupamos el flujo de Roe, los valores de las velocidades de onda se calculan por medio de las ecuaciones (5.2).

Si nodo en la frontera de cama seca tiene seco el lado derecho, $h_l > 0$ y $h_r = 0$, entonces

$$\tilde{\lambda}_1(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u}_l - \sqrt{gh_l}, \quad \tilde{\lambda}_2(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u}_l + 2\sqrt{gh_l}. \quad (5.2a)$$

De otra forma, cuando $h_l = 0$ y $h_r > 0$, entonces

$$\tilde{\lambda}_1(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u}_r - 2\sqrt{gh_r}, \quad \tilde{\lambda}_2(\mathbf{u}_l, \mathbf{u}_r) = \tilde{u}_r + \sqrt{gh_r}. \quad (5.2b)$$

Además de esto, una pequeña capa de agua $\epsilon > 0$ y un caudal $q = 0$ es definida sobre los nodos secos, (Ying et al., 2003). Es decir, al principio del algoritmo, $h = \epsilon$ en los nodos secos y si durante la ejecución resulta que en un lado de un nodo $h > \epsilon$ y del otro $h \leq \epsilon$, entonces se calcula los flujos numéricos de acuerdo a las ecuaciones (5.1) o (5.2). Si en ambos lados la profundidad es menor o igual a ϵ , entonces se establece que los flujos numéricos son iguales a cero, la profundidad del agua se cambia a $h = \epsilon$ y el caudal a $q = 0$.

En la simulación ocupamos el método de RK-TVD y el limitador *minmod* con una malla con 400 elementos, $CFL = 0.01$ y $\epsilon = 10^{-6}$. Los resultados después de 20 segundos son mostrados en la figura 5.17. En ella vemos que la solución concuerda con la exacta, salvo en el frente de la onda que presenta un pequeño desfase. En la figura 5.18 podemos apreciar que el máximo de los errores relativos a lo largo del tiempo de simulación disminuye. El número de CFL no pudo crecer más allá de $1/3$ debido a que necesitamos aumentar la capa de agua ϵ para mantener estabilidad en la solución.

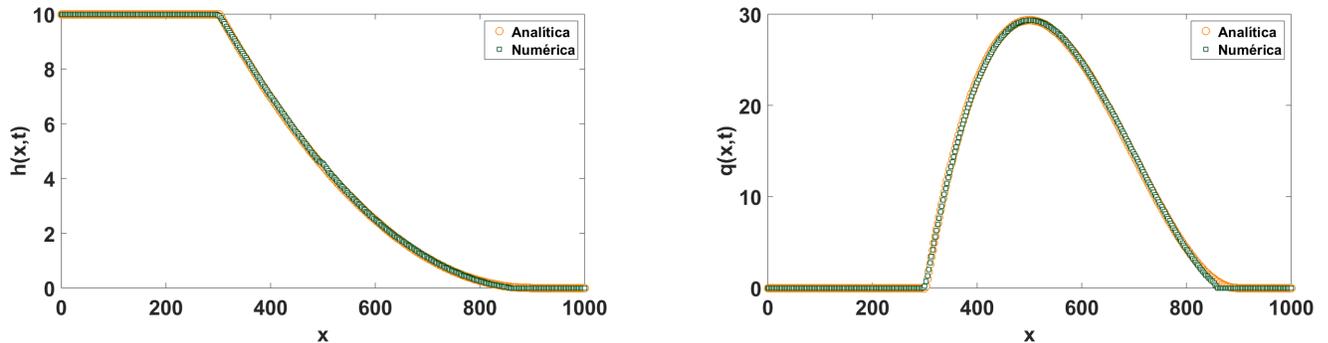


Figura 5.17: Solución para fractura de presa con cama seca. Profundidad $h(x,t)$ (izquierda) y caudal $q(x,t)$ (derecha).

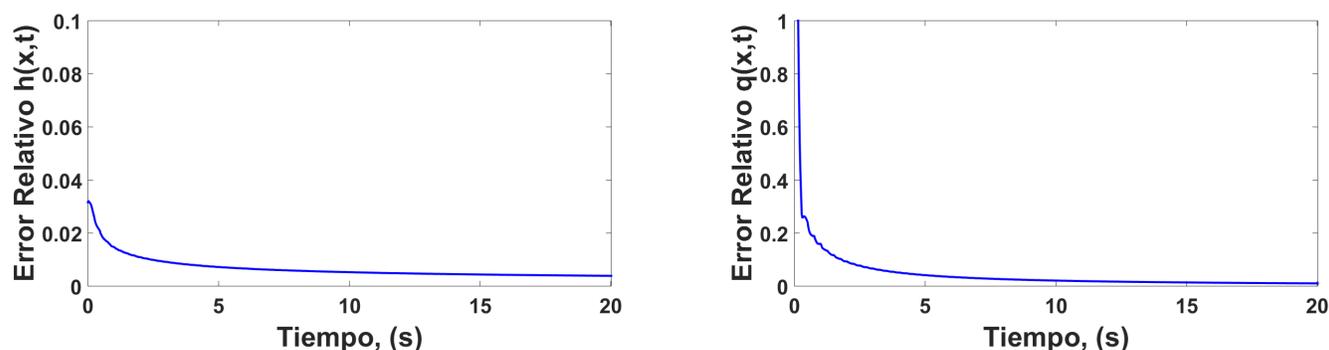


Figura 5.18: Errores relativos para la solución del problema de fractura de presa con cama seca. Profundidad $h(x, t)$ (izquierda) y caudal $q(x, t)$ (derecha).

5.2.3. Fractura de presa en un canal artificial rectangular con fricción sobre lecho seco

En este caso consideramos una fractura sobre un canal artificial de madera horizontal, de 0.096 m de ancho, 0.08 m de altura y 20 m de longitud con una presa colocada a 10 m. El coeficiente de rugosidad de Manning es igual a $0.009 \text{ s}/\text{m}^{1/3}$. La elevación del agua al inicio dentro de la presa es de 0.074 m y con cama seca en el exterior. Se rompe la presa y el flujo es simulado. Una capa de agua $\epsilon = 10^{-4}$ es ocupada.

Al comparar los métodos de Euler hacia adelante y RK-TVD de segundo orden, figuras 5.19 y 5.20 notamos que tiene un comportamiento más suave. Además se requiere aumentar la capa de agua que cuando no hay fricción, ϵ , ya que pierde regularidad en caso contrario.

Cabe mencionar que se esperaba comparar la simulación con datos reales dados en (Schocklitsch, 1917), pero el artículo no fue hallado. Sin embargo, comparando las gráficas con otras presentadas en artículos que ocupan la misma referencia vemos que la simulación es buena.

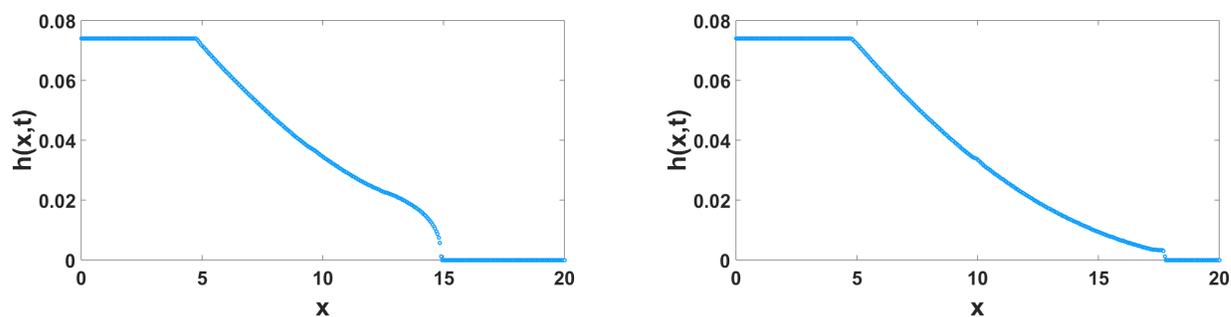


Figura 5.19: Solución para la profundidad fractura de presa con cama seca y fricción. Euler a los 9 segundos (izquierda) y RK-TVD de segundo orden a 10 segundos (derecha).

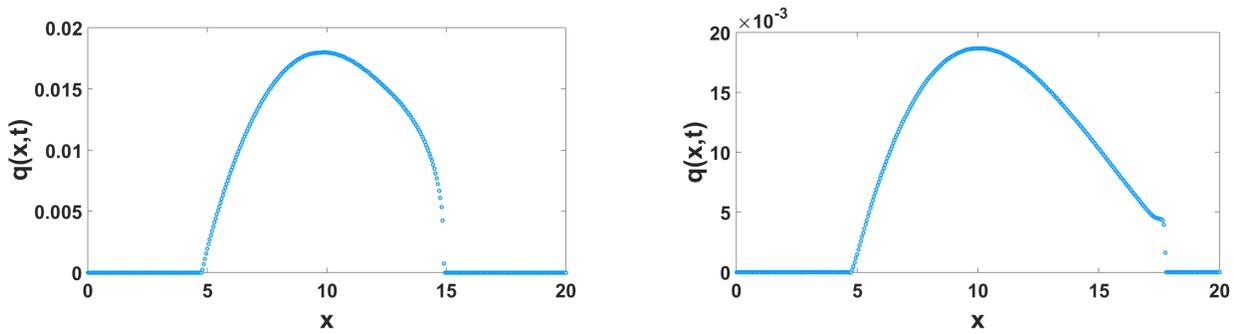


Figura 5.20: Solución para el caudal en la fractura de presa con cama seca y fricción. Euler a los 9 segundos (izquierda) y RK-TVD de segundo orden a 10 segundos (derecha).

5.2.4. Flujo sobre bol parabólico

El siguiente ejemplo también involucra un problema de mojado-seco. En este se aplican el esquema RK-TVD de segundo orden con el limitador *minmod*, además del flujo HLL sobre un bol parabólico. Este escenario consiste en el movimiento de agua dentro del bol. En (Thacker, 1981) se obtiene la solución analítica de este problema. La cama parabólica del canal está determinada por

$$b(x) = h_0 \left(\frac{x^2}{a^2} - 1 \right),$$

donde h_0 y a se muestran en la figura 5.21.

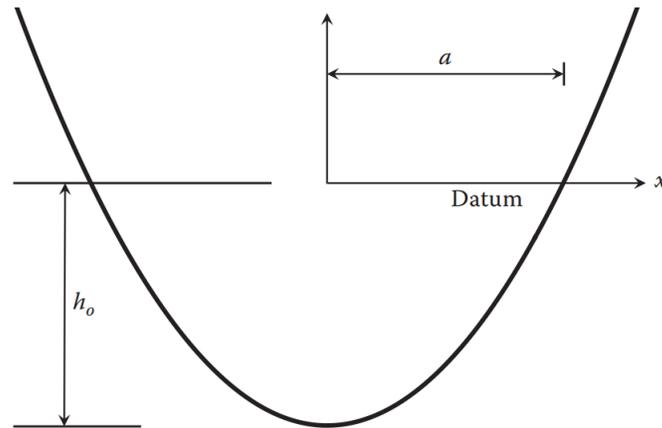


Figura 5.21: Cama para el bol parabólico.

La superficie del agua inicial se representa con la ecuación siguiente:

$$Z(x) = \frac{-2B^2 - 4Bx}{4g} \quad (5.3)$$

donde B es una constante y se tiene agua quieta, esto es, $q = 0$, ver figura 5.22.

La solución analítica para la superficie en cualquier tiempo está dada por

$$Z(x, t) = \frac{-B^2 \cos(2\omega t) - B^2 - 4B\omega \sin(\omega t)x}{4g}, \quad (5.4)$$

y para la velocidad por

$$u(x, t) = B \sin(\omega t), \quad (5.5)$$

con

$$\omega = \frac{\sqrt{2gh_0}}{a}.$$

Claramente, al tratarse de un problema de mojado-seco, es necesario considerar una capa de agua ϵ . Para este caso basta con $\epsilon = 10^{-6}$. En este ejemplo tomamos $h_0 = 10$ m, $a = 600$ m y un tiempo final $T = 230$ s. Los resultados concuerdan con la solución analítica, figuras 5.22 y 5.23.

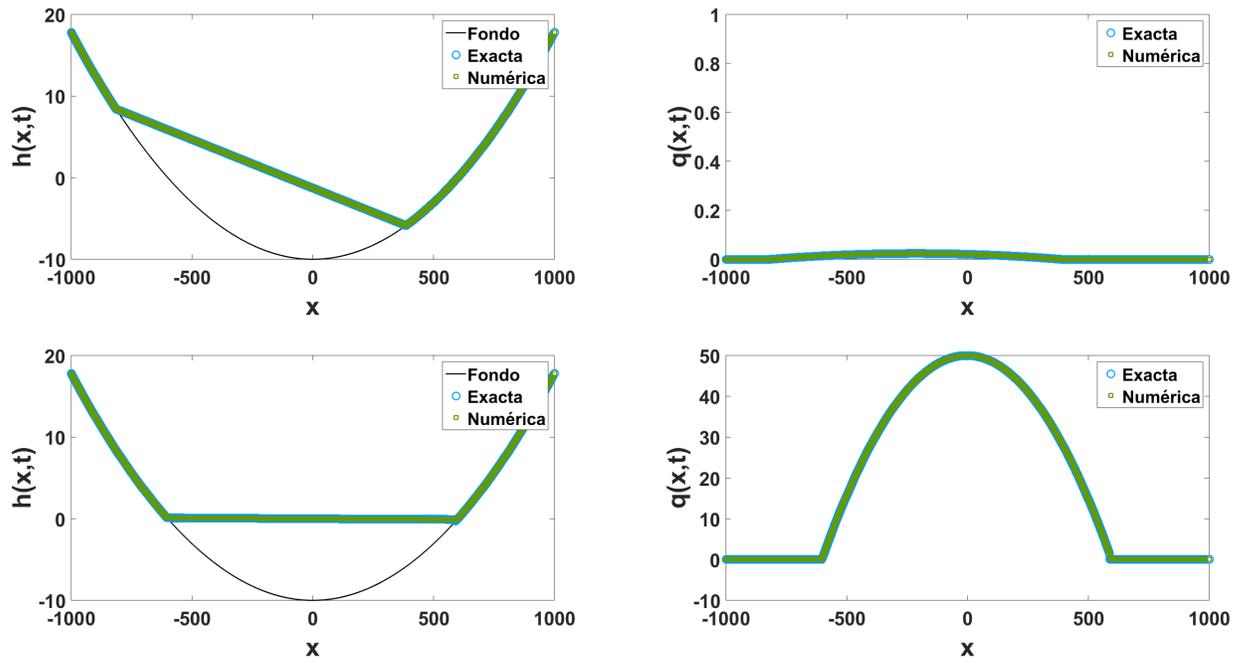


Figura 5.22: Condiciones iniciales para el flujo sobre un bol parabólico (arriba) y después de $t = 67$ s (abajo).

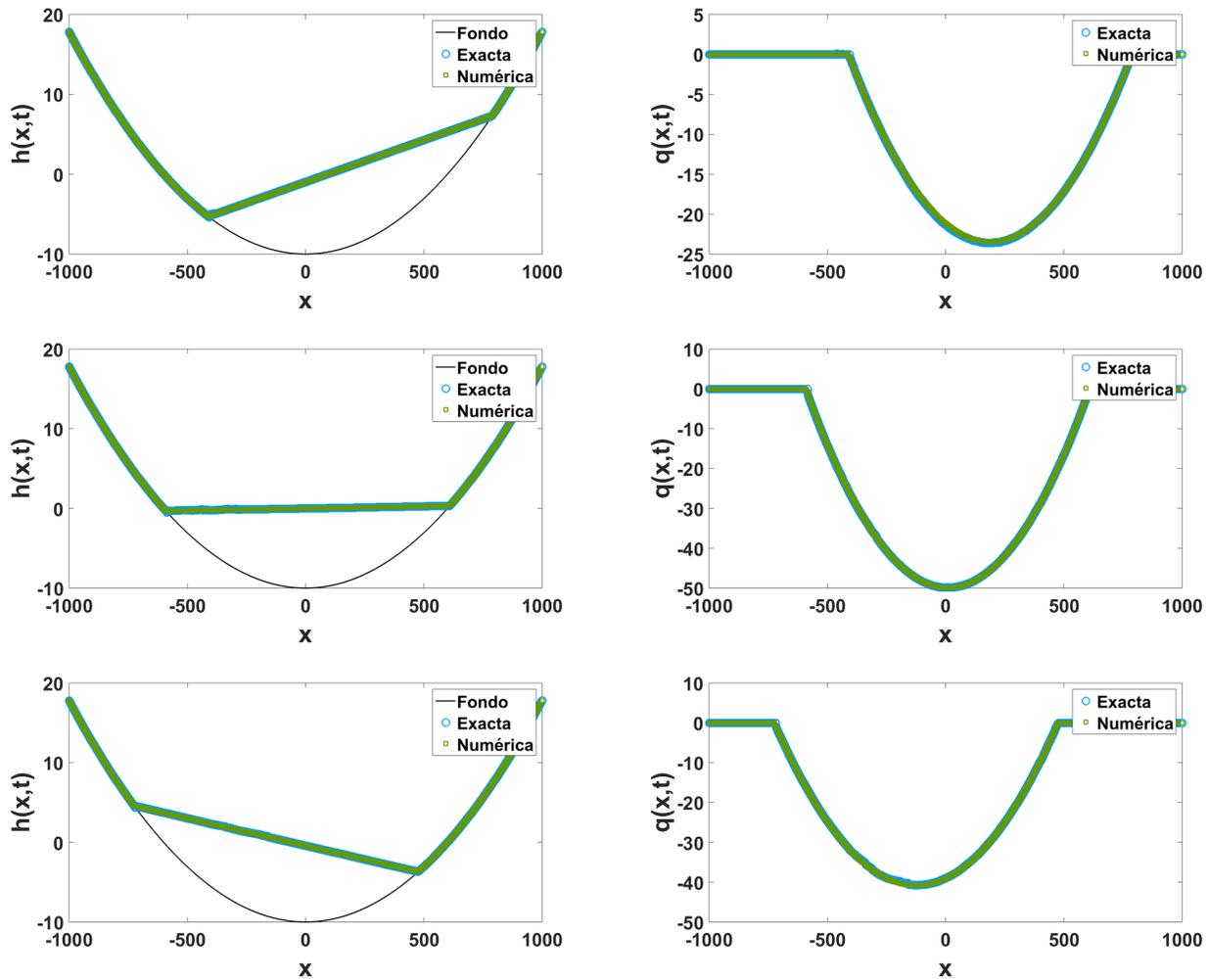


Figura 5.23: Solución sobre un bol parabólico en $t = 135$ s (primera fila), $t = 200$ s (segunda fila) y $t = 230$ s (última fila).

5.2.5. Salto hidráulico

Una simulación de salto hidráulico es simulada como siguiente ejemplo. El canal es de 14 m de longitud y 0.46 m de ancho con una batimetría horizontal. se considera un coeficiente de rugosidad de Manning igual a $0.008 \text{ s/m}^{1/3}$. En la frontera aguas arriba se mantiene la profundidad a 0.031 m y un caudal de $0.118 \text{ m}^2/\text{s}$. En la frontera aguas abajo se establecen condiciones que permiten que el agua salga, es decir, $\frac{\partial q}{\partial x} = 0$. Como condición inicial se utiliza la integración numérica de la siguiente ecuación

$$\frac{dh}{dx} = \frac{S_0 - S_f}{1 - \frac{u^2}{gh}} \quad (5.6)$$

con condición inicial $h(0) = 0.031$, dada en (Gharangik y Chaudhry, 1991). Esta ecuación viene de la idea que al inicio el flujo es estacionario. Implicando que $\frac{\partial q}{\partial t} = 0$ y al sustituir esto último en las ecuaciones de aguas someras resulta la ecuación (5.6).

Así, después de 50 segundos el nivel del agua en la salida sube a 0.265 m. En la figura 5.24 se muestra el estado estacionario de la solución después de 100 segundos con el método de RK-TVD, tomando 100 elementos con un CFL de 0.1 y el flujo HLL.

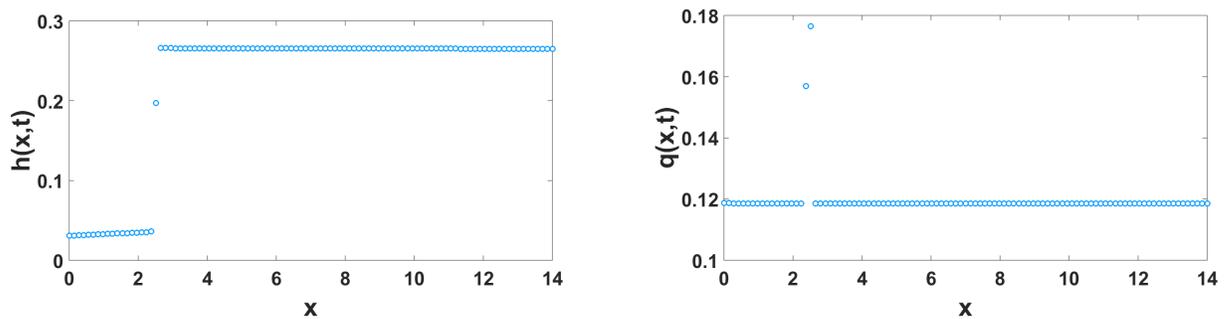


Figura 5.24: Salto hidráulico. Profundidad $h(x, t)$ (izquierda) y caudal $q(x, t)$ (derecha).

5.2.6. Flujo sobre un tope

Varias simulaciones de diferentes regímenes de flujos sobre un tope se presentan. El canal sin fricción es de 1 m de ancho y 25 m de longitud. La cama del canal está definida por

$$b(x) = \begin{cases} 0.2 - 0.05(x - 10)^2, & 8 \leq x \leq 12, \\ 0 & \text{Otro caso.} \end{cases}$$

Se ocupan 100 elementos y un CFL de 0.1. En todas las simulaciones se ocupa el método RK-TVD. El limitador *minmod* puede ser aplicado tanto a la elevación de la superficie del agua, $s = b + h$, como a la profundidad, h . Los resultados después de 200 s se presentan a continuación con el flujo numérico HLL.

Para un flujo subcrítico se considera una superficie de 0.5 m con agua quieta. Entra un flujo de agua con caudal igual a $0.18 \text{ m}^2/\text{s}$. El agua sale del canal a una profundidad de 0.5 m. Al limitar la profundidad, figura 5.25, se obtiene una mejor solución que si limitamos la elevación de la superficie, figura 5.26, particularmente en la solución para el caudal o carga.

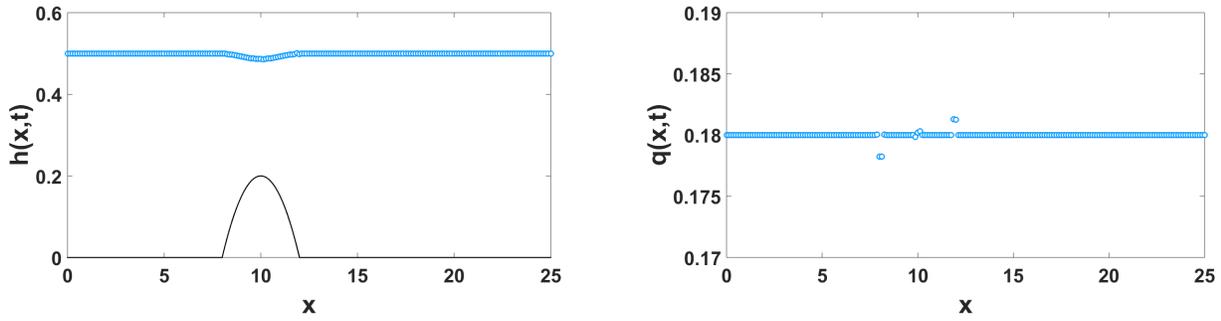


Figura 5.25: Flujo subcrítico limitando la profundidad del agua h .

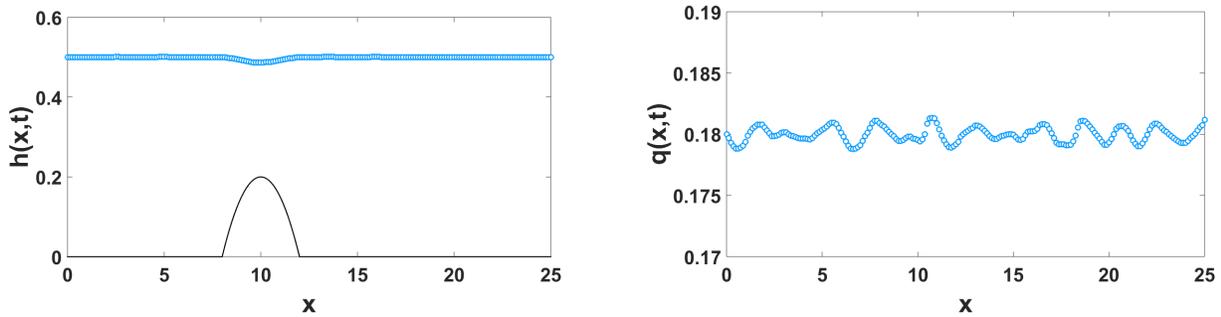


Figura 5.26: Flujo subcrítico limitando la elevación de la superficie del agua $s = b + h$.

Para un flujo transcrito se utiliza una superficie inicial de 0.33 m con agua quieta. Un flujo de agua entra con caudal igual a $0.18 \text{ m}^2/\text{s}$ y la superficie del agua en la salida se mantiene a 0.33 m . El flujo cambia de subcrítico antes del tope a supercrítico sobre el tope y vuelve a ser subcrítico mediante un salto hidráulico. El limitador de superficie es mejor en este caso, como puede observarse en las figuras 5.27 y 5.28.

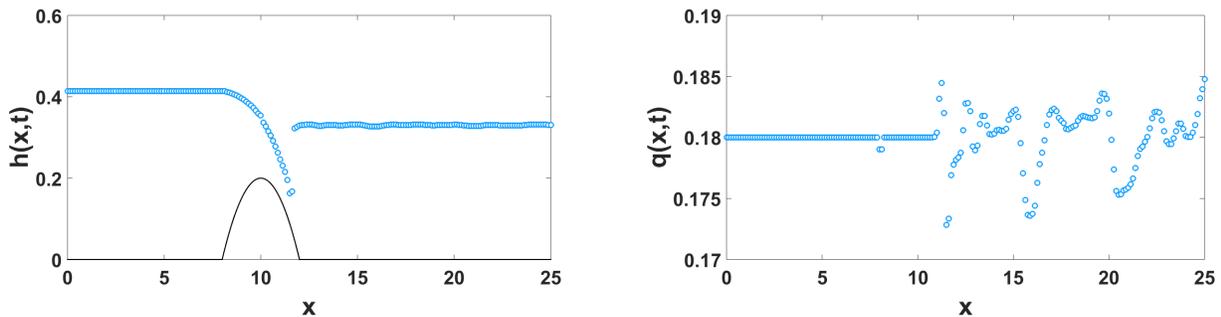


Figura 5.27: Flujo transcrito limitando la profundidad del agua h .

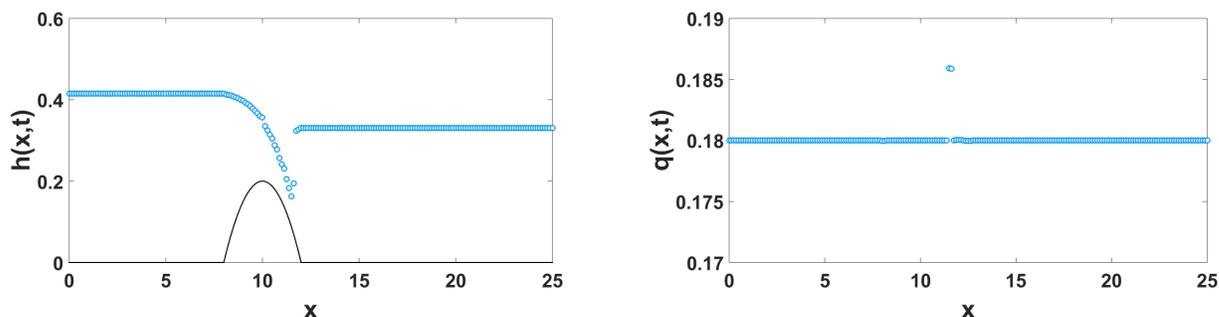


Figura 5.28: Flujo transcrito limitando la superficie del agua $s = b + h$.

Finalmente, un flujo supercrítico es simulado. Las condiciones iniciales y de frontera para este caso son una elevación de la superficie del agua inicial de 2 m con agua quieta. En la entrada del agua la tasa de flujo se mantiene a $25.0567 \text{ m}^2/\text{s}$ con una superficie igual a 2 m. Como en el caso subcrítico, los resultados son mejores al limitar la profundidad, ver figura 5.29.

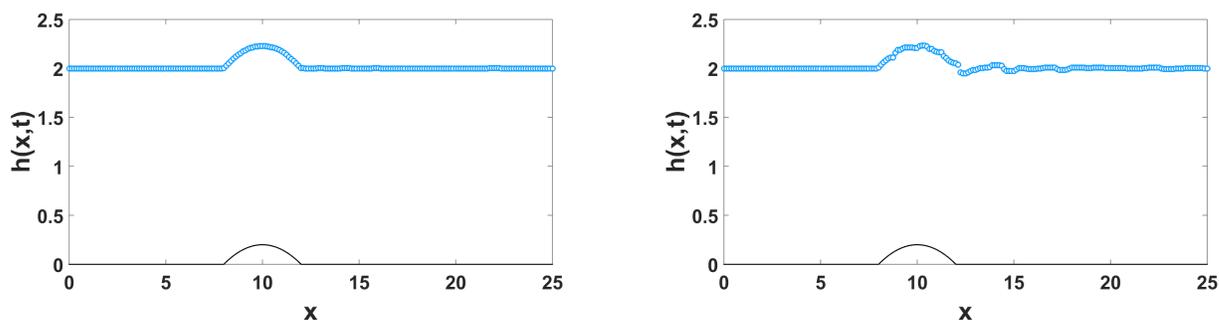


Figura 5.29: Flujo Supercrítico limitando la profundidad (izquierda) o la superficie (derecha).

Con estos resultados concluimos que al considerar el fondo del canal es necesario definir el tipo flujo y donde aplicar el limitador para tener soluciones sin oscilaciones.

5.2.7. Flujo sobre un lecho irregular

En esta prueba se utiliza un flujo transcrito en un canal con cama irregular y sin fricción. La altura de la cama está dada por la tabla 5.2. El método de RK-TVD es utilizado junto con los flujos HLL y de Roe y el limitador *minmod* sobre la superficie.

La superficie inicial es de 16 m con agua quieta. El caudal en la entrada es igual a $50 \text{ m}^2/\text{s}$ y la superficie en la salida se mantiene a 16 m. El flujo dentro del canal cambia de subcrítico a supercrítico y regresa a subcrítico por medio de un salto hidráulico. A partir de los resultados del ejemplo anterior, limitamos la superficie. Además, se tuvo que realizar un refinamiento

x(m)	0	50	100	150	200	250	300	350	400	425
b(m)	0	0	2.5	5	5	3	5	5	7.5	8
x(m)	435	450	470	475	500	505	530	550	565	575
b(m)	9	9	9	9.1	9	9	6	5.5	5.5	5
x(m)	600	650	700	750	800	820	900	950	1000	1500
b(m)	4	3	3	2.3	2	1.2	0.4	0	0	0

Tabla 5.2: Elevación de la cama para cama irregular.

de la malla en la zona más alta de la cama, esto debido a que existe un cambio repentino de pendiente pasando esta zona donde aparecen oscilaciones. Gracias a este refinamiento las oscilaciones pudieron ser reducidas. Los flujos HLL y Roe fueron aplicados obteniendo resultados similares. La simulación después de 3500 s se muestra en las figuras 5.30 y 5.31.

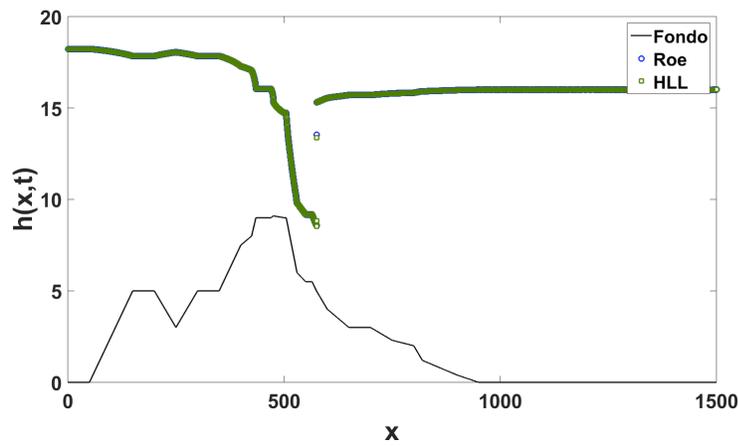


Figura 5.30: Flujo sobre una cama irregular con los flujos numéricos de Roe y HLL.

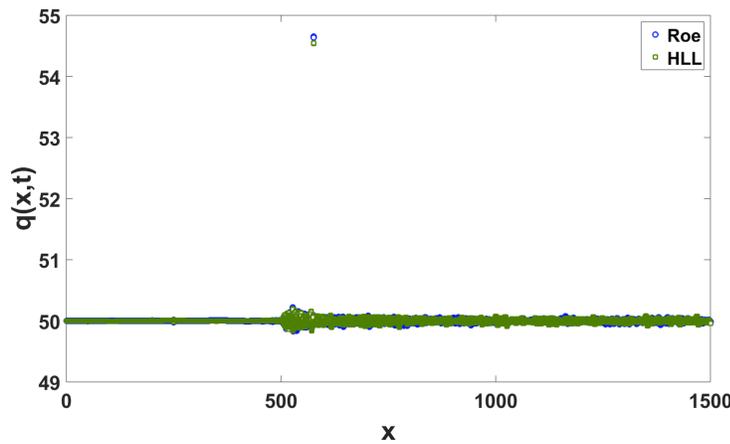


Figura 5.31: Flujo sobre una cama irregular con los flujos numéricos de Roe y HLL.

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

A lo largo de esta tesis se resolvieron numéricamente varios ejemplos para las ecuaciones de aguas someras. Esto con ayuda del método de Galerkin discontinuo para resolver en el espacio y los esquemas de Euler hacia adelante y Runge-Kutta de segundo orden para el tiempo. Este último siendo reescrito para que satisfaga ser un esquema que disminuyan la variación total y así reducir las oscilaciones en la solución y aumentar su precisión, principalmente para ondas de choque. A pesar de esto, oscilaciones aparecen en aquellas zonas cercanas a discontinuidades o con pendientes pronunciadas. Es por esto que se implementaron los limitadores de pendiente *minmod* o TVBM, con los cuales se desea eliminar las oscilaciones. La función *minmod* resultó ser el limitador más adecuado para las ecuaciones de Saint-Venant.

Por otro lado, dependiendo del tipo de escenario a resolver, se tuvieron que realizar consideraciones para lograr una mejor solución. Estas van desde la definición de una capa de agua para los problemas de seco-mojado hasta la necesidad de limitar sobre la superficie o la profundidad del agua en aquellos problemas donde no halla cama plana. Además, se implementaron los flujos numéricos HLL, de Lax-Friedrichs y de Roe como una forma para conectar las soluciones elementales y se concluyó que los flujos HLL y de Roe ofrecen los mejores resultados, tanto en precisión como en la posibilidad de dar pasos de tiempo más grandes.

Pudimos observar que estos esquemas ofrecen buenos resultados, aunque es preferible trabajar en paralelo y con mallas no uniformes. Esto debido a que trabajar secuencialmente aumenta el tiempo de ejecución significativamente debido a la naturaleza local del esquema y la necesidad de tener refinamientos en aquellas zonas donde halla pendientes pronunciadas. Además, los esquemas son condicionalmente estables. Implicando que sea necesario mantener la condición CFL acotada para asegurar soluciones estables.

6.2. Trabajo futuro

Como paso siguiente está la paralelización de este método para lograr un menor tiempo de ejecución. También programar GD en dos y tres dimensiones para trabajar con problemas más reales.

Una posible aplicación de esta metodología a la vida real es resolver problemas de inundaciones. Esto es de importancia vital, ya que podría ayudar a evitar o disminuir los desastres que ocasionan los fenómenos naturales.

Por último, podría aplicarse para resolver ecuaciones diferenciales parciales parabólicas, haciendo las consideraciones pertinentes.

Apéndice A

Ecuaciones hiperbólicas

En este apartado introducimos los conceptos básicos de ecuaciones hiperbólicas que se utilizan para el desarrollo de este trabajo. Empezamos el análisis con una ecuación unidimensional, su clasificación, su formulación débil y el problema de Riemann asociado a la ecuación de Burgers. Posteriormente, aplicamos estas ideas al análisis de sistemas hiperbólicos en una dimensión. Más detalles pueden consultarse en (Mattheij et al., 2005).

A.1. Ecuaciones hiperbólicas escalares de primer orden

Consideremos la expresión

$$a(x, t, u) \frac{\partial u}{\partial t} + b(x, t, u) \frac{\partial u}{\partial x} = c(x, t, u), \tag{A.1}$$
$$a(x, t, u) \neq 0, \quad b(x, t, u) \neq 0.$$

Esta relación es la forma general de las ecuaciones hiperbólicas escalares (en una dimensión) de primer orden.

A.1.1. Ecuaciones lineales y el método de las características

La ecuación (A.1) es llamada lineal cuando a , b y c dependen de x y de t pero no de u . Consideremos $u = \phi(x, t)$ una solución de la ecuación (A.1) y definimos a la superficie $F(x, t, u) := \phi(x, t) - u = 0$. Esta es la superficie integral de la ecuación (A.1) y la denotamos por \mathcal{S} . Además, un vector normal sobre la superficie se denota por $\mathbf{n} = \nabla F = (\phi_x, \phi_t, -1)^T$. Así, para un desplazamiento suficientemente pequeño $d\mathbf{u} = (dx, dt, du)^T$ a lo largo de la superficie, obtenemos

$$\mathbf{n} \cdot d\mathbf{u} = \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial t} dt - du = 0.$$

Al comparar esto con la ecuación (A.1) notamos que para una solución $u = \phi(x, t)$ debe cumplirse lo siguiente sobre la superficie integral \mathcal{S}

$$\begin{pmatrix} a & b \\ dt & dx \end{pmatrix} \begin{pmatrix} \phi_t \\ \phi_x \end{pmatrix} = \begin{pmatrix} c \\ du \end{pmatrix}. \quad (\text{A.2})$$

La unicidad de la solución se obtiene si y sólo si $adx - bdt \neq 0$.

Definimos por $\mathcal{J} = \{x(\sigma), t(\sigma), u(\sigma) | \sigma \in \mathbb{R}\}$, a la curva de valores iniciales. Sea $u \in \mathcal{J}$ y sea \mathcal{C} una curva sobre \mathcal{S} donde $adx - bdt = 0$, figura A.1. Entonces el sistema (A.2) tiene dos posibles resultados, o la ecuación no tiene solución, o existen una infinidad de soluciones. Para este último caso los renglones de la matriz en (A.2) son dependientes sobre la curva \mathcal{C} , llevando a la expresión siguiente:

$$\frac{dt}{a} = \frac{dx}{b} = \frac{du}{c} = \lambda.$$

Al tomar la parametrización $\mathcal{C} = \{(x(s), t(s), u(s))^T | s \in \mathbb{R}\}$, con la condiciones

$$(x(0), t(0), u(0)) \in \mathcal{J}$$

y

$$\frac{dt}{a} = \frac{dx}{b} = \frac{du}{c} = ds,$$

obtenemos el conjunto de ecuaciones diferenciales ordinarias

$$\frac{dt}{ds} = a \quad \frac{dx}{ds} = b \quad \frac{du}{ds} = c, \quad (\text{A.3})$$

con condiciones iniciales $u(0; \sigma) = u(x(\sigma), t(\sigma))$ para $(x(\sigma), t(\sigma)) \in \mathcal{J}'$, donde \mathcal{J}' es la proyección de \mathcal{J} sobre el plano x-t. En la literatura no hay distinción entre \mathcal{J} y \mathcal{J}' , por lo que usaremos \mathcal{J} indistintamente.

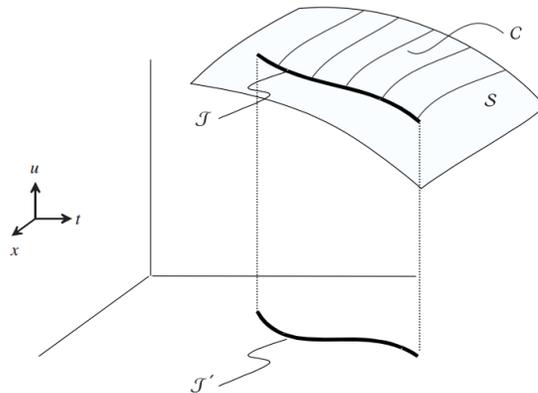


Figura A.1: Superficie integral

Las ecuaciones (A.3) son conocidas como las *ecuaciones características* y \mathcal{C} es una *curva característica*. Observemos que \mathcal{C} es una curva sobre la superficie integral \mathcal{S} y su proyección sobre el plano (x, t) determina la base característica. Luego, para construir \mathcal{S} calculamos la característica que pasa por cada punto sobre \mathcal{J} usando las ecuaciones (A.3). Obteniendo $t = t(s; \sigma)$, $x = x(s; \sigma)$ y $u = u(s; \sigma)$. Si el jacobiano $x_s t_\sigma - x_\sigma t_s \neq 0$, las primeras relaciones pueden invertirse y llegar a $s = s(x, t)$ y $\sigma = \sigma(x, t)$ y al sustituir en la tercer relación resulta $u = u(s; \sigma) = u(s(x, t); \sigma(x, t)) = u(x, t)$.

Podemos hacer una simplificación de estas ecuaciones si parametrizamos la curva \mathcal{C} con t en vez de s y suponiendo que $a \neq 0$. Así,

$$\frac{dx}{dt} = \frac{b}{a}, \quad (\text{A.4a})$$

$$\frac{du}{dt} = \frac{c}{a}. \quad (\text{A.4b})$$

La ecuación (A.4a) determina la posición de la base característica, y la ecuación (A.4b) la solución u como función de t a lo largo de esta base característica. De esta manera, es fácil ver que lo único que se necesita para hallar una solución de (A.1) es resolver (A.4).

En el caso en que la ecuación de primer orden (A.1) sea homogénea y u está dada en algún punto sobre alguna característica \mathcal{C} podemos hallar la solución de la manera siguiente: Sea \mathcal{J} una curva que no sea característica que cruce al menos una vez cada característica, dada por

$$\mathcal{J} = \{(x(\sigma), t(\sigma))^T : \sigma \in I \subset \mathbb{R}\}, \quad (\text{A.5})$$

y denotemos a u sobre \mathcal{J} como

$$u(x(\sigma), t(\sigma)) = v(\sigma), \quad \sigma \in I, \quad (\text{A.6})$$

donde v es una función dada. De esta manera u está completamente determinada por la EDO (A.4) y la condición inicial (A.6).

Ahora, supondremos que a y b son constantes y que \mathcal{J} coincide con el eje x , es decir, podemos escribir $\mathcal{J} = \{(x(\sigma), 0)^T | \sigma \in \mathbb{R}\} = \{(x_0, 0)^T | x_0 \in \mathbb{R}\}$. Entonces, al integrar la ecuación (A.4a) encontramos que la característica que interseca con \mathcal{J} en $(x_0, 0)^T$

$$x - \frac{b}{a}t = x_0,$$

y la solución a lo largo de la característica es

$$u(x, t) = v(x_0) = v\left(x - \frac{b}{a}t\right). \quad (\text{A.7})$$

El valor b/a puede interpretarse como una velocidad a la cual se propaga la solución inicial $v(x_0)$ a lo largo de las características una distancia $x = bt/a$ al tiempo t .

Es importante notar que la solución inicial v no tienen por que ser suave, mucho menos continua. Ver figuras A.2

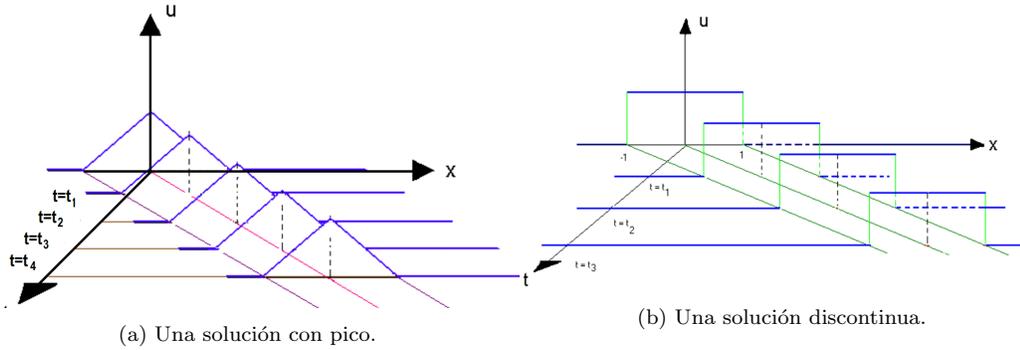


Figura A.2: Diferentes tipos de soluciones propagandose sobre las características.

Cuando $c \neq 0$ encontramos una forma integral de la solución a lo largo de la característica a partir de (A.4b), de la forma

$$u(x, t) = u(x(s_0), t(s_0)) + \int_{t_0}^t \frac{c(x(\tau), \tau)}{a(x(\tau), \tau)} d\tau. \quad (\text{A.8})$$

Si a y b no son constantes, las características son curvas en lugar de líneas rectas, pero podemos proceder de la misma forma para encontrar una solución.

A.1.2. Ecuaciones cuasilineales

La ecuación (A.1) es *cuasilineal* si los coeficientes a , b y c dependen de u . Además, bajo la suposición que $a \neq 0$ podemos llevar esta ecuación a la forma

$$\frac{\partial u}{\partial t} + b(x, t, u) \frac{\partial u}{\partial x} = c(x, t, u). \quad (\text{A.9})$$

Por otro lado, si $b = b(u)$ no depende explícitamente de x y de t , definimos una *función de flujo* $f(u)$ como

$$f(u) := \int_{u_0}^u b(v) dv. \quad (\text{A.10})$$

donde u_0 es un valor de referencia para u . Entonces, podemos reescribir la ecuación (A.9) en su *forma conservativa*

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = c(x, t, u). \quad (\text{A.11})$$

Considerando una vez más el caso homogéneo, $c = 0$, la solución de (A.11) se encuentra aplicando las ecuaciones características (A.4).

$$\frac{dx}{dt} = b(u) = f'(u), \quad (\text{A.12a})$$

$$\frac{du}{dt} = 0. \quad (\text{A.12b})$$

De la relación (A.12a) se sigue que la posición de la característica depende de u en contraste con el caso lineal. En principio se puede calcular u de estas ecuaciones si u está dada en alguna curva inicial \mathcal{J} que cruce las características al menos una vez. Si \mathcal{J} está sobre el eje x , entonces para alguna función v ,

$$u(x, 0) = v(x). \quad (\text{A.13})$$

De la igualdad (A.12b) es posible concluir que u es constante sobre cada característica, y por lo tanto $b(u)$ también lo es. La integración de las ecuaciones (A.12) es sencilla,

$$x - x_0 = b(v(x_0))t, \quad (\text{A.14a})$$

$$u(x, t) = v(x_0) = v(x - b(v(x_0))t). \quad (\text{A.14b})$$

La ecuación (A.14a) describe a x_0 como función de x y t . Sustituyendo esto en la ecuación (A.14b) obtenemos u .

Un ejemplo clásico de este tipo de ecuaciones es la ecuación de Burgers

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0. \quad (\text{A.15})$$

Claramente $b(u) = u$ y se tiene que $u(x, t) = v(x - v(x_0)t)$ a lo largo de las características dadas por $x - x_0 = v(x_0)t$. Como antes, la condición inicial se propaga a lo largo de las curvas características, pero ahora con velocidad dependiente de la posición. Esto puede provocar problemas como vemos en el ejemplo siguiente. Tomemos la ecuación de Burgers con una condición inicial $v(x)$ constante a trozos dada por

$$v(x) := \begin{cases} \alpha & \text{si } x < 0, \\ \beta & \text{si } x \geq 0. \end{cases} \quad (\text{A.16})$$

La característica que atraviesa el punto $(x_0, 0)^T$ está dada por

$$x = x_0 + v(x_0)t, \quad x_0 \in \mathbb{R}.$$

o equivalentemente

$$t = \frac{x - x_0}{v(x_0)}, \quad x_0 \in \mathbb{R},$$

que tiene una pendiente $\frac{dt}{dx} = 1/v(x_0)$. De donde se pueden encontrar dos tipos de soluciones: cuando $\alpha > \beta$ o $\alpha \leq \beta$. Para el primer caso la pendiente de las características que surgen del eje x negativo es más pequeña que la pendiente de las características que emergen del eje x positivo ya que $1/\alpha < 1/\beta$. Esto implicara una solución multivaluada en la intersección de las características, figura A.3a. En el segundo caso, las características de la izquierda tienen una pendiente más grande que las de la derecha, pues $1/\alpha \geq 1/\beta$, llevando a una región en forma de cuña en el plano (x, t) donde la solución no está definida, figura A.3b.

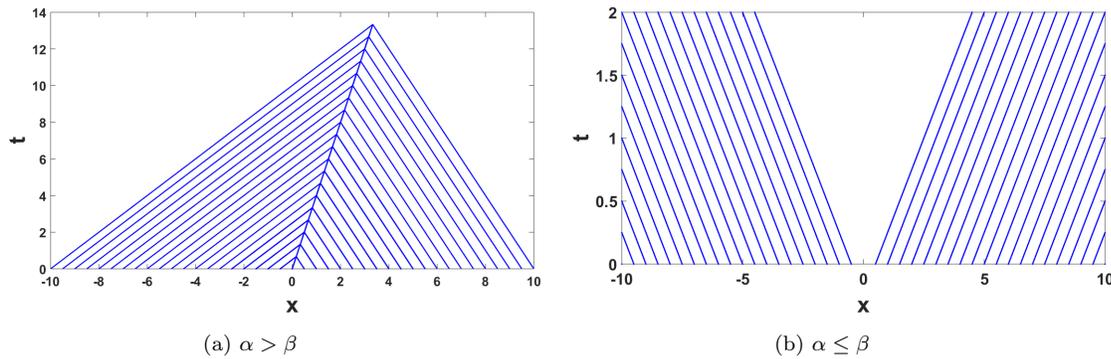


Figura A.3: Características de la ecuación de Burgers.

Pareciera que la discontinuidad de la condición inicial es la causa de este fenómeno, pero es posible tener una condición inicial continua y aun así llegar a una solución que sea multivaluada a partir de cierto tiempo. Por ejemplo, si se establece que la condición inicial para la ecuación de Burgers sea $v(x) = \sin(\pi x)$ para $0 \leq x \leq 1$. Esta función es monótona creciente en el intervalo $(0, 1/2)$ y monótona decreciente en $(1/2, 1)$ y a partir de las ecuaciones (A.14) sabemos que la condición inicial $v(x)$ se propaga a lo largo de las características con velocidad $b(v(x)) = v(x)$. De esta forma, las características que surgen de $(0, 1/2)$ se abren en abanico ocasionando que la condición inicial se expanda. Por otro lado, las características que se originan en $(1/2, 1)$ se aproximan una a otra, compactando la condición inicial en este intervalo, ver figura A.4. Esto significa que la parte izquierda de la solución se estira hacia la derecha, sobrepasando a la parte derecha, conduciendo a un perfil más abrupto conforme pasa el tiempo, figura A.5.

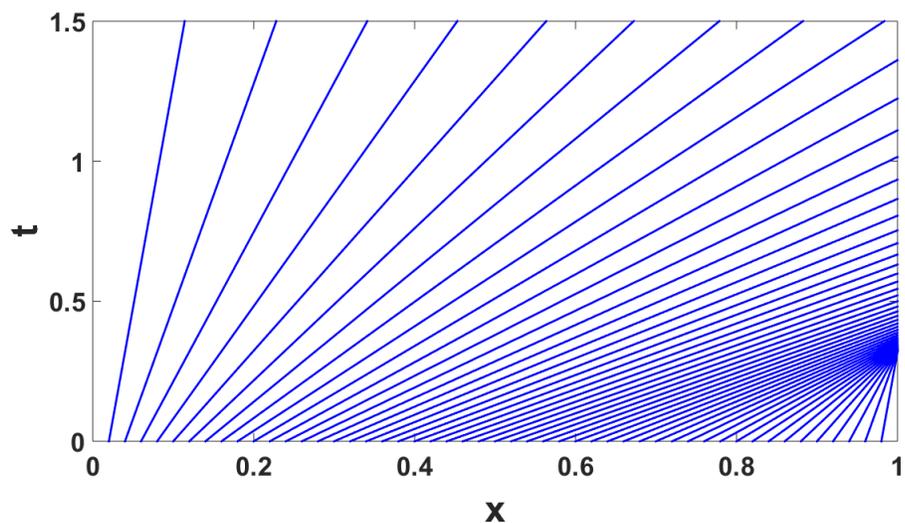


Figura A.4: Características de la ecuación de Burgers con condición inicial $v(x) = \sin(\pi x)$

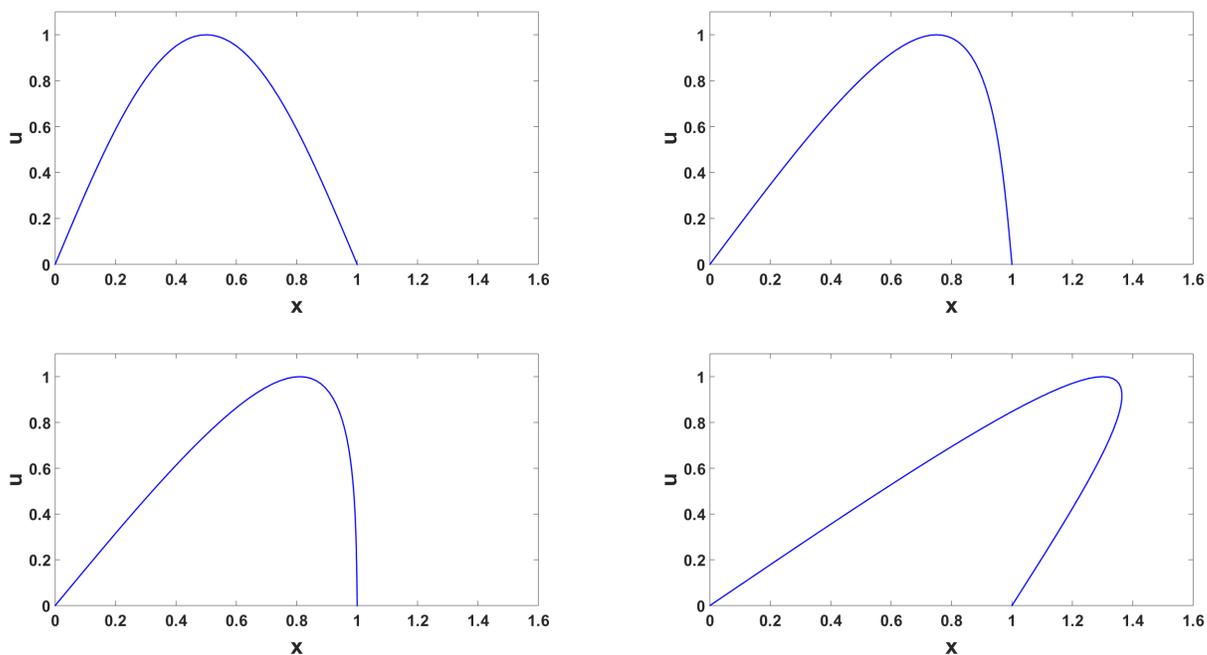


Figura A.5: Soluciones de la ecuación de burgers con condición inicial $v(x) = \sin(\pi x)$ en los tiempos $t = 0s$, $t = 0.25s$, $t = 0.30s$ y $0.80s$.

En conclusión, el comportamiento de la solución no sólo depende de las condiciones iniciales, sino también de la función de flujo que describe el problema.

A.2. Formulación débil de ecuaciones escalares de primer orden

La definición de solución débil que se usa está basada en teoría de distribuciones. Sea el espacio de funciones de prueba

$$\begin{aligned} D &= \{ \phi \in \mathcal{C}^1(\mathbb{R} \times [0, \infty)) \mid \phi \text{ tiene soporte compacto para cualquier } t \}, \\ &= \mathcal{C}_0^1(\mathbb{R} \times [0, \infty)). \end{aligned}$$

Al multiplicar la ecuación (A.11), con $c = 0$, por una función de prueba $\phi \in D$,

$$\frac{\partial u}{\partial t} \phi + \frac{\partial f(u)}{\partial x} \phi = 0 \phi,$$

luego tomamos la integral sobre $\mathbb{R} \times [0, \infty)$

$$\int_{-\infty}^{\infty} \int_0^{\infty} \left[\frac{\partial u}{\partial t} \phi + \frac{\partial f(u)}{\partial x} \phi \right] dt dx = \int_{-\infty}^{\infty} \int_0^{\infty} 0 \phi dt dx,$$

separando y aplicando el método de integración por partes obtenemos

$$\int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial u}{\partial t} \phi dt dx + \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial f(u)}{\partial x} \phi dt dx = 0,$$

$$\begin{aligned} & \int_{-\infty}^{\infty} \left[\lim_{b \rightarrow \infty} (u\phi)(x, b) - (u\phi)(x, 0) - \int_0^{\infty} u \frac{\partial \phi}{\partial t} dt \right] dx \\ & + \int_0^{\infty} \left[\lim_{a \rightarrow \infty} (f(u)\phi)(a, t) - \lim_{a \rightarrow -\infty} (f(u)\phi)(a, t) - \int_{-\infty}^{\infty} f(u) \frac{\partial \phi}{\partial x} dx \right] dt = 0, \\ & - \int_{-\infty}^{\infty} u(x, 0) \phi(x, 0) dx = \int_{-\infty}^{\infty} \int_0^{\infty} \left[u \frac{\partial \phi}{\partial t} + f(u) \frac{\partial \phi}{\partial x} \right] dt dx. \end{aligned} \quad (\text{A.18})$$

Donde se ha utilizado que $\phi(x, t) \rightarrow 0$ cuando $|x| + |t| \rightarrow \infty$. Así, podemos obtener la definición siguiente:

Definición A.1. Una función $u(x, t)$ se llama *solución débil* de la ley de conservación (A.11) si (A.18) se cumple para toda función de prueba $\phi \in D$

Notemos que si u satisface (A.11) automáticamente es solución débil, lo contrario sólo pasará si u es $\mathcal{C}^1(\mathbb{R} \times [0, \infty))$.

Observemos además que la relación (A.18) permite soluciones discontinuas. Sin embargo, no toda solución discontinua puede ser solución de (A.11), como muestra el teorema siguiente.

Teorema A.2.1. *Sea u una solución suave a trozos de (A.11) que tiene una discontinuidad*

a través de una curva $\epsilon : x = x(t)$. Entonces u debe satisfacer la condición

$$[f(u)]_-^+ = s [u]_-^+,$$

con $[v]_-^+ := v(x(t)^+, t) - v(x(t)^-, t)$ el salto de la variable v a través de la discontinuidad ϵ y s es la velocidad de la discontinuidad.

Podemos concluir que la velocidad de propagación s de una discontinuidad es

$$s = \frac{f(u_r) - f(u_l)}{u_r - u_l}, \quad (\text{A.19})$$

donde $u_r := u(x(t)^+, t)$ y $u_l := u(x(t)^-, t)$ son los límites laterales derecho e izquierdo, respectivamente, en la discontinuidad.

La relación (A.19) se llama la *condición de salto de Rankine-Hugoniot*, y puede reescribirse como:

$$s = \frac{1}{u_r - u_l} \int_{u_l}^{u_r} b(v) dv,$$

que permite interpretar a s como el promedio de la velocidad de advección en el intervalo entre u_l y u_r .

Cuando se estudian soluciones débiles debe imponerse la condición de salto de Rankine-Hugoniot a la EDP, ya que esta representa la ley de conservación física correcta que atraviesa la discontinuidad.

A.3. El problema de Riemann

A continuación analizamos un problema clásico, el problema de Riemann. Este es un problema de valor inicial con condición inicial constante a trozos y una sola discontinuidad, el cual tiene la siguiente forma:

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0, \quad (\text{A.20a})$$

$$u(x, 0) := \begin{cases} u_l & \text{si } x < 0, \\ u_r & \text{si } x > 0. \end{cases} \quad (\text{A.20b})$$

Para resolver esta ecuación buscamos soluciones autosimilares. Esto se logra convirtiendo la EDP con variables (x, t) en una EDO con variable η . Así, sea $u(x, t) = \hat{u}(x/t) = \hat{u}(\eta)$, entonces, al derivar y sustituir en la ecuación sucede que

$$\frac{\partial \hat{u}}{\partial t} + \frac{\partial f(\hat{u})}{\partial x} = 0,$$

$$\begin{aligned}
\frac{d\hat{u}}{d\eta} \frac{\partial \eta}{\partial t} + \frac{df(\hat{u})}{d\hat{u}} \frac{d\hat{u}}{d\eta} \frac{\partial \eta}{\partial x} &= 0, \\
-\frac{\partial \hat{u}}{\partial \eta} \frac{x}{t^2} + b(\hat{u}) \frac{\partial \hat{u}}{\partial \eta} \frac{1}{t} &= 0, \\
\frac{1}{t} [b(\hat{u}) - \eta] \frac{d\hat{u}}{d\eta} &= 0, \\
[b(\hat{u}) - \eta] \frac{d\hat{u}}{d\eta} &= 0.
\end{aligned}$$

dando lugar a los casos siguientes:

1. $\frac{d\hat{u}}{d\eta} = 0$, que implica una solución constante sobre el rayo $x/t = \eta$, cuya forma es $\tilde{u}(\eta) = \tilde{u}(x/t) = u(x, 0)$, donde $u(x, 0)$ sea continua.
2. $b(\hat{u}) - \eta = 0$, es decir, $b(\hat{u}) = \eta$, implicando que, siempre que b sea invertible, la solución es $\hat{u} = b^{-1}(\eta)$, donde la condición inicial tenga discontinuidades.

Como vimos al principio de este apéndice, las condiciones iniciales del problema (A.20) se propagan a lo largo de las características con velocidad $b(u) = \frac{df}{du}$. Basado en esto, el análisis se hace en los casos $b(u_l) > b(u_r)$ y $b(u_l) < b(u_r)$.

Las características tienen la forma $x = b(v_0)t + x_0$, o equivalentemente, $t = \frac{x - x_0}{b(v_0)}$ y $u(x, t) = v(x_0) = v(x - tb(v_0))$.

Caso 1: $b(u_l) > b(u_r)$

En este caso, las características que surgen del eje negativo tendrán una pendiente más pequeña que las que surgen del lado positivo. Dando así una intersección de las características, que podría convertirse en una solución multivaluada. En vez de esto, la solución será discontinua y está dada por la relación

$$u(x, t) = \begin{cases} u_l & \text{si } x/t < s, \\ u_r & \text{si } x/t > s. \end{cases} \quad (\text{A.21})$$

donde s está dada por la condición de Rankine-Hugoniot (A.19). Podemos interpretar esta solución como la discontinuidad inicial viajando a velocidad s y se llama *onda de choque*. La velocidad s es la *velocidad de choque*.

Como ejemplo, la ecuación de Burgers con una condición inicial de salto para mostrar el comportamiento de las características.

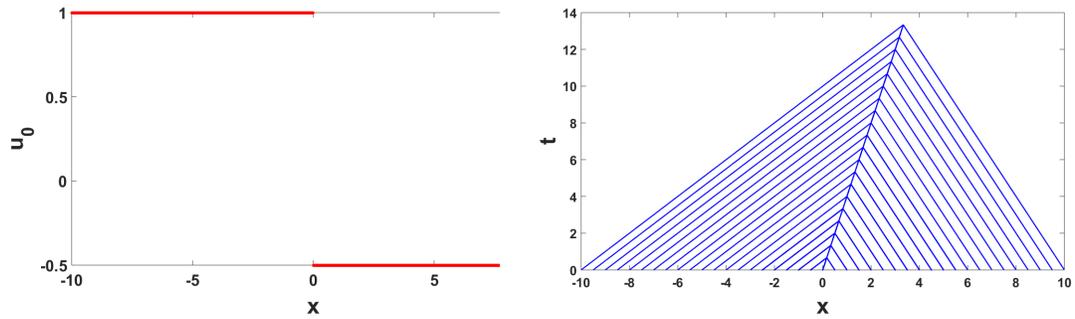


Figura A.6: Características del problema de Riemann con $u_l > u_r$.

Caso 2: $b(u_l) < b(u_r)$

Ahora las pendientes que surgen en el lado positivo serán más pequeñas que las que salen del lado negativo. Así las características se separaran y sobre la región en forma de cuña la solución será continua. Esta solución está descrita por la relación

$$u(x, t) = \begin{cases} u_l & \text{si } x/t < b(u_l), \\ \tilde{u}(x/t) & \text{si } b(u_l) < x/t < b(u_r), \\ u_r & \text{si } b(u_r) < x/t, \end{cases} \quad (\text{A.22})$$

donde $\tilde{u}(x/t)$ es una solución de similitud de (A.20a). Esta solución es conocida como *onda de rarefacción* y es continua a pesar del hecho de que tiene una discontinuidad en la condición inicial. La solución $u(x, t)$ consiste de estados constantes u_l y u_r unidos por la solución intermedia $\tilde{u}(x/t)$.

De nuevo, se toma como ejemplo a la ecuación de Burgers cambiando la condición de salto.

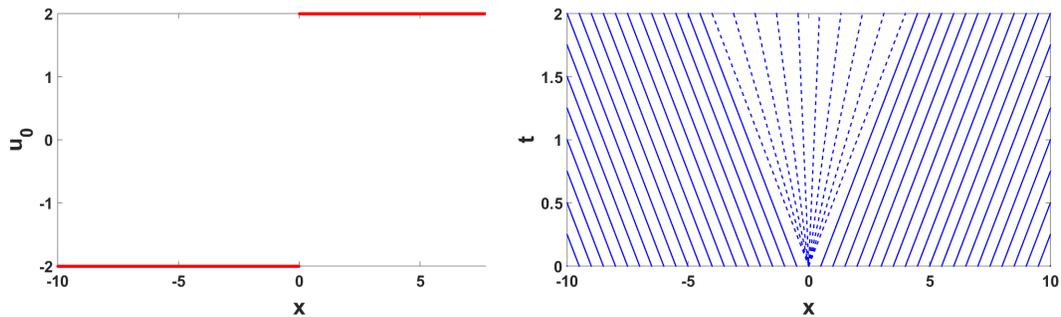


Figura A.7: Características del problema de Riemann con $u_l < u_r$.

Es posible demostrar que la solución (A.21) también es solución del problema de Riemann cuando $b(u_l) < b(u_r)$. Este tipo de solución, llamada *choque de expansión*, podría llevar a la no unicidad de la solución en el segundo caso. Sin embargo, el choque de expansión es físicamente incorrecto y debe evitarse. Para hacerlo se utiliza condiciones de entropía como

la *condiciones de entropía de Lax*, dada por la definición siguiente:

Definición A.2. Una solución discontinua de (A.20a) que tiene una función de flujo convexa (cóncava) satisface la condición de entropía si

$$b(u_l) > s > b(u_r), \quad (b(u_l) < s < b(u_r),) \quad (\text{A.23})$$

con s la velocidad de la discontinuidad definida en las condiciones de Rankine-Hugoniot (A.19).

La solución que cumpla con esta condición es única. Si la función de flujo de la ecuación no es convexa ni cóncava se utiliza la siguiente definición.

Definición A.3 (*Condición de entropía de Oleinik*). Una solución débil de (A.20a) es solución de entropía si todas las discontinuidades, las cuales se propagan con la velocidad s dada en (A.19), satisfacen para todo u entre u_l y u_r :

$$\frac{f(u) - f(u_l)}{u - u_l} \geq s \geq \frac{f(u) - f(u_r)}{u - u_r}. \quad (\text{A.24})$$

En conclusión, hemos obtenido una manera de resolver el problema de Riemann para una ecuación diferencial parcial en una dimensión.

A.4. Sistemas de primer orden

Ahora extendemos el discurso anterior a sistemas de primer orden en los casos lineal y cuasilineal.

A.4.1. Sistemas lineales

Considérese el siguiente sistema de ecuaciones

$$\mathbf{A} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{B} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{c}, \quad (\text{A.25})$$

con \mathbf{A} y \mathbf{B} matrices de $m \times m$ constantes y \mathbf{c} un vector de \mathbb{R}^m que puede depender de x , t y \mathbf{u} . Este sistema puede desacoplarse cuando la matriz \mathbf{A} es no singular con ayuda de los valores y los vectores propios derechos de \mathbf{BA}^{-1} . Sea \mathbf{S} la matriz cuyas columnas son los vectores propios derechos de \mathbf{BA}^{-1} y $\mathbf{\Lambda}$ la matriz diagonal cuyas entradas son los valores propios de \mathbf{BA}^{-1} entonces

$$\begin{aligned} \mathbf{S}^{-1} \mathbf{BA}^{-1} \mathbf{S} &= \mathbf{\Lambda}, \\ \hat{\mathbf{u}}(x, t) &= \mathbf{S}^{-1} \mathbf{A} \mathbf{u}(x, t). \end{aligned}$$

Con el que puede llegarse al sistema desacoplado

$$\frac{\partial \hat{\mathbf{u}}}{\partial t} + \mathbf{\Lambda} \frac{\partial \hat{\mathbf{u}}}{\partial x} = \mathbf{S}^{-1} \mathbf{c},$$

o equivalentemente a

$$\frac{\partial \hat{u}_k}{\partial t} + \lambda_k \frac{\partial \hat{u}_k}{\partial x} = \hat{c}_k, \quad k = 1, 2, \dots, m.$$

Así, cada uno de estas ecuaciones forma una curva característica \mathcal{C}_k correspondiente al valor propio λ_k y al vector propio \mathbf{s}_k . Las ecuaciones características serán

$$\frac{dx}{dt} = \lambda_k, \quad \frac{d\hat{u}_k}{dt} = \hat{c}_k,$$

Llegando a la siguiente definición

Definición A.4. El sistema lineal (A.25) será hiperbólico si la matriz \mathbf{BA}^{-1} tiene m valores propios reales y \mathbf{S} es una matriz no singular.

Esto puede ser generalizado al caso en que \mathbf{A} es singular. Para ello se usan los vectores propios generalizados izquierdos. Esto es, encontrar un vector propio izquierdo $\mathbf{t}^T \neq \mathbf{0}$ y su correspondiente valor propio generalizado λ tal que

$$\lambda \mathbf{t}^T \mathbf{A} = \mathbf{t}^T \mathbf{B},$$

donde el $\det(\lambda \mathbf{A} - \mathbf{B})$ debe ser cero. La matriz $\lambda \mathbf{A} - \mathbf{B}$ se llama matriz lápiz y si $\det(\lambda \mathbf{A} - \mathbf{B}) = 0, \forall \lambda \in \mathbf{C}$, entonces la matriz lápiz es singular.

Asumiendo que la matriz lápiz es regular, esto es, existen m valores propios reales $\lambda_i, i = 1, 2, \dots, m$ tales que $\det(\lambda \mathbf{A} - \mathbf{B}) \neq \mathbf{0}$ y que la matriz es completa, es decir, existe un conjunto de m vectores propios izquierdos $\mathbf{t}_i^T, i = 1, 2, \dots, m$, es posible expresar el problema de hallar los valores y vectores propios generalizados en la forma siguiente:

Encontrar λ_i y \mathbf{t}_i^T , con $i = 1, 2, \dots, m$, tales que:

$$\mathbf{TB} = \mathbf{\Lambda TA}, \tag{A.26}$$

con

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m), \quad \mathbf{T} = \begin{pmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2^T \\ \vdots \\ \mathbf{t}_m^T \end{pmatrix}.$$

Al multiplicar por la izquierda la ecuación (A.25) por la matriz \mathbf{T} y sustituir en la ecuación (A.26) resulta

$$\mathbf{TA} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{\Lambda TA} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{Tc}.$$

Como antes, $\hat{\mathbf{u}}(x, t) = \mathbf{T}\mathbf{A}\mathbf{u}(x, t)$ y se obtiene el sistema desacoplado

$$\frac{\partial \hat{\mathbf{u}}}{\partial t} + \Lambda \frac{\partial \hat{\mathbf{u}}}{\partial x} = \hat{\mathbf{c}} = \mathbf{T}\mathbf{c}, \quad (\text{A.27a})$$

o, escrita en términos de cada componente

$$\frac{\partial \hat{u}_k}{\partial t} + \lambda_k \frac{\partial \hat{u}_k}{\partial x} = \hat{c}_k, \quad k = 1, 2, \dots, m. \quad (\text{A.27b})$$

Entonces, es posible generalizar la definición de sistema hiperbólico.

Definición A.5 (*Sistema hiperbólico*). El sistema (A.25) será hiperbólico si existen una matriz diagonal real Λ y una matriz no singular \mathbf{T} tales que $\mathbf{T}\mathbf{B} = \Lambda\mathbf{T}\mathbf{A}$.

A.4.2. Sistemas cuasilineales

Ahora consideramos el caso en el que las matrices \mathbf{A} y \mathbf{B} dependan explícitamente de x , t y \mathbf{u}

$$\mathbf{A}(x, t, \mathbf{u}) \frac{\partial \mathbf{u}}{\partial t} + \mathbf{B}(x, t, \mathbf{u}) \frac{\partial \mathbf{u}}{\partial x} = \mathbf{c}(x, t, \mathbf{u}).$$

Suponiendo que \mathbf{A} es no singular, la ecuación anterior tiene la forma

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{B}(x, t, \mathbf{u}) \frac{\partial \mathbf{u}}{\partial x} = \mathbf{c}(x, t, \mathbf{u}). \quad (\text{A.28})$$

Definición A.6. Diremos que el sistema (A.28) es hiperbólico en (x, t, \mathbf{u}) si existe una matriz diagonal real $\Lambda(x, t, \mathbf{u})$ y una matriz no singular $\mathbf{T}(x, t, \mathbf{u})$ que consisten de los valores y vectores propios generalizados, tales que

$$\mathbf{T}(x, t, \mathbf{u})\mathbf{B}(x, t, \mathbf{u}) = \Lambda(x, t, \mathbf{u})\mathbf{T}(x, t, \mathbf{u}). \quad (\text{A.29})$$

Es claro ver que ahora la hiperbolicidad de la ecuación (A.28) dependerá de x , t y \mathbf{u} .

Al multiplicar por la izquierda la ecuación (A.28) por \mathbf{T} se llega a

$$\mathbf{T} \frac{\partial \mathbf{u}}{\partial t} + \Lambda \mathbf{T} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{T}\mathbf{c} = \tilde{\mathbf{c}}. \quad (\text{A.30})$$

Lo siguiente es introducir las variables características $\tilde{\mathbf{u}}$ por medio de la *ecuación diferencial de Pfaff*

$$d\tilde{\mathbf{u}} = \mathbf{T}d\mathbf{u} \quad (\text{A.31})$$

donde d es una diferencial arbitraria. Esta ecuación expresa la diferencial de $\tilde{\mathbf{u}}$ como una combinación lineal de la diferencial de \mathbf{u} . Si la ecuación (A.31) es integrable, (de Matemáticas de la Universidad de Extremadura, 2013), el sistema (A.30) puede ser reducido a

$$\frac{\partial \tilde{\mathbf{u}}}{\partial t} + \mathbf{\Lambda} \frac{\partial \tilde{\mathbf{u}}}{\partial x} = \tilde{\mathbf{c}}. \quad (\text{A.32})$$

Luego, el sistema (A.28) puede ser diagonalizado, pero sigue estando acoplado debido a que la matriz $\mathbf{\Lambda}$ depende de la variable \mathbf{u} .

Por otro lado, si se multiplica el vector propio izquierdo \mathbf{t}_k^T de \mathbf{B} por la ecuación (A.28) y se considera la identidad (A.29) obtenemos

$$\mathbf{t}_k^T \frac{\partial \mathbf{u}}{\partial t} + \mathbf{t}_k^T \mathbf{B} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{t}_k^T \left(\frac{\partial \mathbf{u}}{\partial t} + \lambda_k \frac{\partial \mathbf{u}}{\partial x} \right) = \mathbf{t}_k^T \mathbf{c} = \tilde{c}_k. \quad (\text{A.33})$$

Se desea que esta ecuación sea equivalente a la EDO de la forma

$$\mathbf{t}_k^T \frac{d\mathbf{u}}{ds} = \tilde{c}_k, \quad (\text{A.34})$$

la cual se satisface sobre alguna curva $\mathcal{K} = \{(x(s), t(s)) | s \in I \subset \mathbb{R}\}$. Al usar la regla de la cadena para $\frac{d\mathbf{u}}{ds}$ resulta

$$\frac{d\mathbf{u}}{ds} = \frac{\partial \mathbf{u}}{\partial t} \frac{dt}{ds} + \frac{\partial \mathbf{u}}{\partial x} \frac{dx}{ds}.$$

Al sustituir la relación anterior en (A.34) y comparar el resultado con la ecuación (A.33) se concluye que

$$\frac{dt}{ds} = 1, \quad \frac{dx}{ds} = \lambda_k. \quad (\text{A.35a})$$

Además, de la ecuación de Pfaff se tiene que

$$\frac{d\tilde{u}_k}{ds} = \mathbf{t}_k^T \frac{d\mathbf{u}}{ds} = \tilde{c}_k. \quad (\text{A.35b})$$

Luego se tiene que

$$\frac{dt}{ds} = 1, \quad \frac{dx}{ds} = \lambda_k, \quad \frac{d\tilde{u}_k}{ds} = \tilde{c}_k \quad (\text{A.36})$$

son las ecuaciones características asociadas a la ecuación hiperbólica

$$\frac{\partial \tilde{u}_k}{\partial t} + \lambda_k \frac{\partial \tilde{u}_k}{\partial x} = \tilde{c}_k.$$

Por ello se concluye que la curva \mathcal{K} es la curva característica \mathcal{C}_k correspondiente al valor propio λ_k . La ecuación (A.34) es la *forma normal o forma característica*. Las variables $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_m)$ son las variables de Riemann. Si $\mathbf{c} = \mathbf{0}$, las variables \tilde{u}_k serán constantes a lo largo de la respectiva característica \mathcal{C}_k y se dice que son las *invariantes de Riemann*.

En los casos en los que la matriz \mathbf{B} y el vector \mathbf{c} dependan únicamente de \mathbf{u} , el sistema (A.28) puede reescribirse en forma conservativa como:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = \mathbf{c}(\mathbf{u}). \quad (\text{A.37})$$

donde la función de flujo $\mathbf{f}(\mathbf{u})$ está relacionada a la matriz $\mathbf{B}(\mathbf{u})$ de la forma

$$\mathbf{B}(\mathbf{u}) = \frac{\partial \mathbf{f}(\mathbf{u})}{\partial \mathbf{u}}.$$

Es decir, $\mathbf{B}(\mathbf{u})$ es la matriz jacobiana de $\mathbf{f}(\mathbf{u})$. La ecuación será hiperbólica si $\mathbf{B}(\mathbf{u})$ tiene m valores propios reales y es diagonalizable.

A.4.3. Método de las características

La forma normal (A.34), junto con las ecuaciones características (A.36) conducen a un método de solución.

Consideremos el caso lineal para el sistema (A.25), con $m = 2$,

$$\begin{aligned} a_{11} \frac{\partial u_1}{\partial t} + a_{12} \frac{\partial u_2}{\partial t} + b_{11} \frac{\partial u_1}{\partial x} + b_{12} \frac{\partial u_2}{\partial x} &= c_1, \\ a_{21} \frac{\partial u_1}{\partial t} + a_{22} \frac{\partial u_2}{\partial t} + b_{21} \frac{\partial u_1}{\partial x} + b_{22} \frac{\partial u_2}{\partial x} &= c_2. \end{aligned}$$

Dado que el sistema es hiperbólico, existen dos valores propios λ_v y λ_w . Sea $\mathbf{v}^T = (v_1, v_2)$ el valor propio izquierdo asociado a λ_v , entonces, $\mathbf{v}^T \mathbf{B} = \lambda_v \mathbf{v}^T \mathbf{A}$. Sean $\tilde{\mathbf{v}}^T = \mathbf{v}^T \mathbf{A} = (\tilde{v}_1, \tilde{v}_2)$ y $\tilde{c}_1 = \mathbf{v}^T \mathbf{c}$.

Se define ξ como la variable independiente a lo largo de la característica, la cual es descrita por las ecuaciones

$$\frac{dt}{d\xi} = 1, \quad \frac{dx}{d\xi} = \lambda_v \quad (\text{A.39})$$

Esto implica que

$$\begin{aligned} \mathbf{v}^T \mathbf{A} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{v}^T \mathbf{B} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{v}^T \mathbf{c} &\iff \tilde{\mathbf{v}}^T \frac{\partial \mathbf{u}}{\partial t} + \lambda_v \tilde{\mathbf{v}}^T \frac{\partial \mathbf{u}}{\partial x} = \tilde{c}_1, \\ &\iff \tilde{v}_1 \left(\frac{\partial u_1}{\partial t} + \lambda_v \frac{\partial u_1}{\partial x} \right) + \tilde{v}_2 \left(\frac{\partial u_2}{\partial t} + \lambda_v \frac{\partial u_2}{\partial x} \right) = \tilde{c}_1, \\ &\iff \tilde{v}_1 \frac{\partial u_1}{\partial \xi} + \tilde{v}_2 \frac{\partial u_2}{\partial \xi} = \tilde{c}_1 \iff \left(\frac{\partial u_1}{\partial \xi}, \frac{\partial u_2}{\partial \xi} \right) (\tilde{v}_1, \tilde{v}_2)^T = \tilde{c}_1, \\ &\iff \frac{\partial \mathbf{u}^T}{\partial \xi} \mathbf{v} = \tilde{c}_1. \end{aligned}$$

Es decir,

$$D_{\mathbf{v}}\mathbf{u} = \tilde{c}_1 \quad (\text{A.40})$$

Esto indica que la derivada de \mathbf{u} en la dirección del valor propio izquierdo \mathbf{v} es constante. Del mismo modo, para λ_w existe el vector propio izquierdo $\mathbf{w} = (w_1, w_2)$ y sea $\tilde{\mathbf{w}}^T = \mathbf{w}^T \mathbf{A} = (\tilde{w}_1, \tilde{w}_2)$ y $\tilde{c}_2 = \mathbf{w}^T \mathbf{c}$. Al definirse η como la variable independiente correspondiente a la característica que satisface

$$\frac{dt}{d\eta} = 1, \quad \frac{dx}{d\eta} = \lambda_w \quad (\text{A.41})$$

entonces

$$D_{\mathbf{w}}\mathbf{u} = \tilde{c}_2 \quad (\text{A.42})$$

Las ecuaciones (A.39) a la (A.42) determinan un sistema completo para hallar las características y las soluciones sobre ellas.

A.5. Formulación débil de sistemas de primer orden

Ahora se define la solución débil para el sistema hiperbólico (A.37). Esta es una generalización de la solución débil vista en el apartado (A.2).

Sea ϕ que pertenece al espacio de funciones de prueba D^m , con $D = \mathcal{C}_0^1(\mathbb{R} \times [0, \infty))$. Al hacer el producto interior entre ϕ y la ecuación (A.37) e integrar sobre $\mathbb{R} \times [0, \infty)$ usando las fórmulas de Green se llega a

$$\int_0^\infty \int_{-\infty}^\infty \left(\mathbf{u} \cdot \frac{\partial \phi}{\partial t} + \mathbf{f}(\mathbf{u}) \cdot \frac{\partial \phi}{\partial x} \right) dx dt = - \int_{-\infty}^\infty \mathbf{u}(x, 0) \cdot \phi(x, 0) dx - \int_0^\infty \int_{-\infty}^\infty \mathbf{c} \cdot \phi dx dt \quad (\text{A.43})$$

Definición A.7. Una función $\mathbf{u}(x, t)$ es una solución débil del sistema (A.37) si la relación (A.43) se satisface para cualquier función de prueba $\phi \in D^m$.

Al igual que en el caso escalar, una solución débil puede ser suave a trozos, donde las partes suaves están separadas por discontinuidades. Para que esta solución sea admisible debe satisfacer la condición de Rankine-Hugoniot, dada por el teorema siguiente:

Teorema A.5.1. Sea $\mathbf{u}(x, t)$ una solución suave a trozos de la ecuación (A.37), que tiene una discontinuidad a través de la curva $\epsilon := x = x(t)$. Entonces \mathbf{u} satisface la condición

$$[\mathbf{f}(\mathbf{u})]_\pm^\pm = s[\mathbf{u}]_\pm^\pm \quad (\text{A.44})$$

donde $[\mathbf{v}]_\pm^\pm := \mathbf{v}(x(t)^\pm, t) - \mathbf{v}(x(t)^\mp, t)$ es el salto de \mathbf{v} a través de ϵ y s es la velocidad de ϵ .

La condición de Rankine-Hugoniot (A.44) es necesaria para la construcción de soluciones de ondas de choque.

A.5.1. El problema de Riemann

En este apartado derivamos las soluciones elementales del problema de Riemann asociado a la ecuación (A.37). Por simplicidad, $\mathbf{c} = 0$. De esta forma, la forma canónica del problema de Riemann se escribe como

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0, \quad (\text{A.45a})$$

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_l, & x < 0, \\ \mathbf{u}_r, & x > 0. \end{cases} \quad (\text{A.45b})$$

Al igual que en el caso escalar, cualquier solución de (A.45) es una solución autosimilar de la forma $\mathbf{u}(x, t) = \tilde{\mathbf{u}}(x/t)$.

Primero analizamos el caso lineal, es decir, cuando $\mathbf{f}(\mathbf{u}) = \mathbf{B}\mathbf{u}$ con \mathbf{B} una matriz constante. Dado que los vectores propios derechos, $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$ de la matriz \mathbf{B} son linealmente independientes, los vectores \mathbf{u}_l y \mathbf{u}_r se pueden expresar como

$$\mathbf{u}_l = \sum_{k=1}^m \alpha_k \mathbf{s}_k, \quad \mathbf{u}_r = \sum_{k=1}^m \beta_k \mathbf{s}_k. \quad (\text{A.46})$$

Por otro lado, la condición inicial también puede representarse usando los valores propios derechos como

$$\mathbf{u}(x, 0) = \mathbf{S}\tilde{\mathbf{u}}(x, 0) = \sum_{k=1}^m \tilde{u}_k(x, 0) \mathbf{s}_k. \quad (\text{A.47})$$

Al comparar la ecuación (A.46) con (A.47) concluimos

$$\tilde{u}_k(x, 0) = \begin{cases} \alpha_k & x < 0 \\ \beta_k & x > 0 \end{cases} \quad (\text{A.48})$$

Dado que los valores propios λ_k son constantes, se puede encontrar la solución \tilde{u}_k utilizando la ecuación (A.27b) y las condiciones (A.48) para llegar a

$$\tilde{u}_k(x, t) = \tilde{u}_k(x - \lambda_k t, 0) = \begin{cases} \alpha_k, & x/t < \lambda_k, \\ \beta_k, & x/t > \lambda_k. \end{cases} \quad (\text{A.49})$$

Entonces, la solución es

$$\mathbf{u}(x, t) = \mathbf{S}\tilde{\mathbf{u}} = \sum_{k=1}^m \tilde{u}_k(x, t) \mathbf{s}_k = \sum_{x/t < \lambda_k} \alpha_k \mathbf{s}_k + \sum_{x/t > \lambda_k} \beta_k \mathbf{s}_k \quad (\text{A.50})$$

Luego, la solución \mathbf{u} es constante a trozos debido a que la discontinuidad inicial en $x = 0$ se propaga a lo largo de todas las características. Los parches o secciones constantes de \mathbf{u} en

el plano (x, t) son separadas por las características. Por ejemplo, si se tiene un sistema de 3×3 , cuyos valores propios satisfacen $\lambda_1 < 0$ y $0 < \lambda_2 < \lambda_3$, entonces la solución en la región I, figura A.8, tiene la forma $\mathbf{u}(x, t) = \alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 + \alpha_3 \mathbf{s}_3$. Esto se interpreta con ayuda de la relación (A.49), pensando que en esta región los rayos $x/t < \lambda_k, \forall k$. Es decir, los estados constantes $(\alpha_1, \alpha_2, \alpha_3)$ son los que se propagan en esta región. Del mismo modo, la solución en las demás regiones puede verse como los estados constantes iniciales moviéndose sobre dichas regiones dependiendo de la ecuación (A.49).

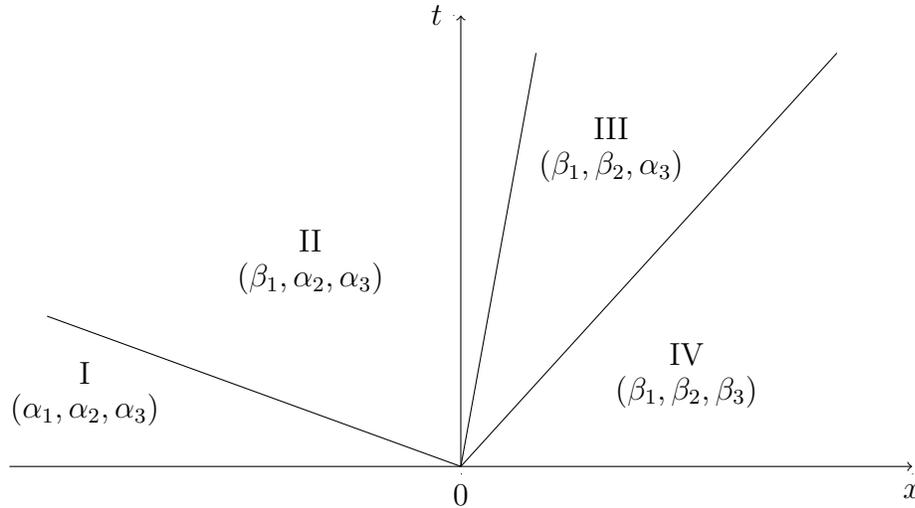


Figura A.8: Soluciones autosimilares del problema de Riemann para un sistema lineal de 3×3 .

Ahora se considera el caso cuasilineal. Debemos notar que la matriz \mathbf{B} depende de \mathbf{u} y por lo tanto también los valores y vectores propios derechos. Además, la búsqueda de la solución se hace por casos, dependiendo de que los vectores propios satisfagan alguna de las siguientes definiciones.

Definición A.8. Un vector propio $\mathbf{s}_k(\mathbf{u})$, con valor propio $\lambda_k(\mathbf{u})$, es genuinamente no lineal si

$$\left(\frac{\partial \lambda_k}{\partial \mathbf{u}}(\mathbf{u}), \mathbf{s}_k(\mathbf{u}) \right) \neq 0, \quad \forall \mathbf{u}, \quad (\text{A.51})$$

donde

$$\frac{\partial}{\partial \mathbf{u}} := \left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}, \dots, \frac{\partial}{\partial u_m} \right).$$

Por otra parte, diremos que $\mathbf{s}_k(\mathbf{u})$ es linealmente degenerado si

$$\left(\frac{\partial \lambda_k}{\partial \mathbf{u}}(\mathbf{u}), \mathbf{s}_k(\mathbf{u}) \right) = 0, \quad \forall \mathbf{u}. \quad (\text{A.52})$$

A continuación, presentamos los tres posibles casos para la solución del problema de Riemann para una ecuación cuasilineal: onda de rarefacción, discontinuidad de contacto y onda de choque.

Caso 1. Onda de rarefacción.

Supongamos que $\lambda_k(\mathbf{u}_l) < \lambda_k(\mathbf{u}_r)$ y $\mathbf{s}_k(\mathbf{u})$ es genuinamente no lineal. Entonces podemos normalizar \mathbf{s}_k de tal forma que

$$\left(\frac{\partial \lambda_k}{\partial \mathbf{u}}(\mathbf{u}), \mathbf{s}_k(\mathbf{u}) \right) = 1, \quad \forall \mathbf{u}.$$

Entonces, para un vector de estados izquierdo, \mathbf{u}_l , podemos asociar el problema de valor inicial

$$\frac{d\tilde{\mathbf{u}}(\eta)}{d\eta} = \mathbf{s}_k(\tilde{\mathbf{u}}(\eta)), \quad \eta > 0, \quad (\text{A.53a})$$

$$\tilde{\mathbf{u}}(0) = \mathbf{u}_l, \quad (\text{A.53b})$$

el cual define una curva integral que es tangente a \mathbf{s}_k en todas partes. Sea $\mathbf{u}_r = \tilde{\mathbf{u}}(\eta_r)$, para algún $\eta_r > 0$. Tomando la derivada

$$\begin{aligned} \frac{d}{d\eta}(\lambda_k(\tilde{\mathbf{u}}(\eta))) &= \frac{\partial \lambda_k}{\partial \mathbf{u}}(\tilde{\mathbf{u}}(\eta))^T \frac{d\tilde{\mathbf{u}}(\eta)}{d\eta} \\ &= \left(\frac{\partial \lambda_k}{\partial \mathbf{u}}(\tilde{\mathbf{u}}(\eta)), \mathbf{s}_k(\tilde{\mathbf{u}}(\eta)) \right) = 1 \end{aligned}$$

Entonces, obtenemos

$$\lambda_k(\tilde{\mathbf{u}}(\eta)) = \eta + \lambda_k(\tilde{\mathbf{u}}_l), \quad \lambda_k(\tilde{\mathbf{u}}_r) = \eta_r + \lambda_k(\tilde{\mathbf{u}}_l) > \lambda_k(\tilde{\mathbf{u}}_l).$$

Así, la curva $\lambda_k(\tilde{\mathbf{u}}(\eta))$ une a los estados $\tilde{\mathbf{u}}_l$ y $\tilde{\mathbf{u}}_r$. Además, podemos ver que $\tilde{\mathbf{u}}(\eta)$ es una solución autosimilar como en el caso escalar. La pregunta ahora es saber cual es el parámetro η en el que la solución es $\tilde{\mathbf{u}}(\eta)$. Para responder esto, tomemos $\lambda_k(\mathbf{u}_l) < x/t < \lambda_k(\mathbf{u}_r)$ o equivalentemente $0 < x/t - \lambda_k(\mathbf{u}_l) < \lambda_k(\mathbf{u}_r) - \lambda_k(\mathbf{u}_l)$. Así, $\eta := x/t - \lambda_k(\mathbf{u}_l)$. Luego

$$\lambda_k(\mathbf{u}(x, t)) = \lambda(\tilde{\mathbf{u}}(x/t - \lambda_k(\mathbf{u}_l))) = x/t - \lambda_k(\mathbf{u}_l) + \lambda_k(\mathbf{u}_l) = x/t.$$

Además

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} &= \frac{\partial \mathbf{u}}{\partial t} + \mathbf{B}(\mathbf{u}) \frac{\partial \mathbf{u}}{\partial x} \\ &= \frac{-x}{t^2} \mathbf{s}_k(\tilde{\mathbf{u}}(\eta)) + \frac{1}{t} \mathbf{B}(\tilde{\mathbf{u}}(\eta)) \mathbf{s}_k(\tilde{\mathbf{u}}(\eta)) \\ &= \frac{1}{t} \left(-\frac{x}{t} + \lambda_k(\tilde{\mathbf{u}}(\eta)) \right) \mathbf{s}_k(\tilde{\mathbf{u}}(\eta)) = \mathbf{0}, \end{aligned}$$

así $\tilde{\mathbf{u}}(\eta)$ es solución de la ecuación (A.45) y podemos escribir la solución para este caso como

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_l & x/t < \lambda_k(\mathbf{u}_l) \\ \tilde{\mathbf{u}}(x/t - \lambda_k(\mathbf{u}_l)) & \lambda_k(\mathbf{u}_l) < x/t < \lambda_k(\mathbf{u}_r) \\ \mathbf{u}_r & x/t > \lambda_k(\mathbf{u}_r) \end{cases} \quad (\text{A.54})$$

conocida como una k -onda de rarefacción.

Para calcular las k -ondas de rarefacción a partir del problema de valor inicial (A.53), es útil usar las k -invariantes de Riemann, las cuales están definidas como sigue.

Definición A.9. Una k -invariante de Riemann del problema de Riemann (A.45) es una función continuamente diferenciable $w_k : \mathbb{R}^m \rightarrow \mathbb{R}$ tal que

$$\left(\frac{\partial w_k}{\partial \mathbf{u}}(\mathbf{u}), \mathbf{s}_k(\mathbf{u}) \right) = 0, \quad \forall \mathbf{u} \quad (\text{A.55})$$

Podemos obtener de esta definición que una k -invariante de Riemann es constante a lo largo de la curva integral de (A.53). Este resultado es ocupado para calcular las k -ondas de rarefacción para el problema de aguas someras.

Caso 2. Discontinuidad de contacto.

Cuando \mathbf{s}_k es linealmente degenerado, se tiene que $\lambda_k(\mathbf{u})$ es constante a lo largo de las curvas integrales que son tangentes a \mathbf{s}_k .

Así, $\lambda_k(\tilde{\mathbf{u}}_l(\eta)) = \lambda_k(\tilde{\mathbf{u}}_l) = \lambda_k(\tilde{\mathbf{u}}_r)$ para $\eta \in [0, \eta_r]$. Podemos demostrar que la solución es

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_l & x/t < \lambda_k(\mathbf{u}_l) \\ \mathbf{u}_r & x/t > \lambda_k(\mathbf{u}_l) \end{cases} \quad (\text{A.56})$$

llamada discontinuidad de contacto. Para lograrlo, debemos probar que (A.56) satisface las condiciones de Rankine-Hugoniot (A.44). Como $\lambda_k(\tilde{\mathbf{u}}(\eta))$ es constante tenemos

$$\frac{d}{d\eta} (\mathbf{f}(\tilde{\mathbf{u}}(\eta)) - \lambda_k(\tilde{\mathbf{u}}(\eta))\tilde{\mathbf{u}}(\eta)) = \mathbf{B}(\tilde{\mathbf{u}}(\eta)) \frac{d\tilde{\mathbf{u}}}{d\eta} - \lambda_k(\tilde{\mathbf{u}}(\eta)) \frac{d\tilde{\mathbf{u}}}{d\eta} \quad (\text{A.57})$$

$$= (\mathbf{B}(\tilde{\mathbf{u}}(\eta)) - \lambda_k(\tilde{\mathbf{u}}(\eta))\mathbf{I}) \frac{d\tilde{\mathbf{u}}}{d\eta} = \mathbf{0}. \quad (\text{A.58})$$

Por lo que $\mathbf{f}(\tilde{\mathbf{u}}(\eta)) - \lambda_k(\tilde{\mathbf{u}}(\eta))\tilde{\mathbf{u}}(\eta) = \mathbf{C} \in \mathbb{R}^m$. En particular, $\mathbf{f}(\tilde{\mathbf{u}}_r) - \lambda_k(\tilde{\mathbf{u}}_r)\tilde{\mathbf{u}}_r = \mathbf{f}(\tilde{\mathbf{u}}_l) - \lambda_k(\tilde{\mathbf{u}}_l)\tilde{\mathbf{u}}_l$, y las condición de Rankine-Hugoniot se satisface con $s = \lambda_k(\tilde{\mathbf{u}}_l)$.

Caso 3. Onda de choque.

Finalmente, supongamos que $\lambda_k(\mathbf{u}_l) > \lambda_k(\mathbf{u}_r)$ y \mathbf{s}_k es genuinamente no lineal. Entonces, al igual que en la versión escalar, las características se intersecan, pudiendo dar una solución multivaluada. En lugar de esto, la solución es discontinua y está dada por

$$\mathbf{u}(x, t) = \begin{cases} \mathbf{u}_l & x/t < s \\ \mathbf{u}_r & x/t > s \end{cases} \quad (\text{A.59})$$

donde s es la velocidad de discontinuidad definida en la condición de salto de Rankine-Hugoniot (A.44). A esta solución se le llama k -onda de choque. Como en el caso escalar, es

necesario determinar si la solución tiene un sentido físico correcto. Con este fin se utilizan las condiciones de entropía, dadas por

Definición A.10. *Condición de entropía de Lax.* La k -onda de choque (A.59) satisface la condición de entropía si se satisfacen las siguientes desigualdades, con s la velocidad de choque dada en (A.44).

$$\lambda_{k-1}(\mathbf{u}_l) < s < \lambda_k(\mathbf{u}_l) \quad (\text{A.60a})$$

$$\lambda_k(\mathbf{u}_r) < s < \lambda_{k+1}(\mathbf{u}_r) \quad (\text{A.60b})$$

con s la velocidad de propagación de la discontinuidad dada por la ecuación (A.44).

La solución general del problema de Riemann involucra las soluciones elementales descritas arriba, como describe el siguiente teorema.

Teorema A.5.2. *Asumiendo que el sistema del problema de Riemann (A.45) es hiperbólico y que cada vector propio de la matriz jacobiana de $\mathbf{f}(\mathbf{u})$ es o bien genuinamente no lineal o linealmente degenerado. Entonces para cualquier $\mathbf{u}_l \in \mathbb{R}^m$ existe una vecindad \mathcal{N} de \mathbf{u}_l tal que el problema de Riemann (A.45) tiene una solución única si $\mathbf{u}_r \in \mathcal{N}$. Esta solución consiste de a lo más $m + 1$ estados constantes separados por ondas de choque, de rarefacción o de contacto.*

Apéndice B

Condiciones de frontera

En este apartado damos las nociones sobre las condiciones de frontera para las ecuaciones de aguas someras. Primero hacemos un paréntesis para hablar de condiciones de frontera en ecuaciones hiperbólicas, para luego aplicar estas ideas al caso de las aguas poco profundas.

B.1. Condiciones de frontera: Caso lineal

Las condiciones iniciales son muy importantes para resolver cualquier ecuación diferencial. Además, sabemos que en problemas hiperbólicos estas condiciones se propagan a lo largo de las curvas características. Implicando que para un problema con dominio infinito basta con dar condiciones iniciales en las variables características para encontrar una solución. Cuando el problema está restringido a una sección del espacio es importante proporcionar condiciones de frontera y saber de donde surgen las características para que el problema esté bien planteado.

Como primer ejemplo y debido a que es el caso más sencillo, empezamos con el problema lineal homogéneo n -dimensional.

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{B} \frac{\partial \mathbf{u}}{\partial x} = \mathbf{0}, \quad x \in (0, 1), \quad t > 0, \quad (\text{B.1a})$$

$$\mathbf{u}(x, 0) = \mathbf{v}(x), \quad x \in (0, 1), \quad (\text{B.1b})$$

$$\mathbf{C}_l \mathbf{u}(0, t) = \mathbf{g}_l(t), \quad \mathbf{C}_r \mathbf{u}(1, t) = \mathbf{g}_r(t), \quad t > 0, \quad (\text{B.1c})$$

donde la matriz \mathbf{B} tiene n valores propios reales,

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-p} < 0 < \lambda_{n-p+1} \leq \dots \leq \lambda_n,$$

con p el número de valores propios positivos. Además, \mathbf{C}_l es una matriz constante de tamaño $n_l \times n$ y $\mathbf{g}_l \in \mathbb{R}^{n_l}$. Estos establecen n_l condiciones de frontera en $x = 0$. De la misma forma, \mathbf{C}_r es una matriz constante de $n_r \times n$ y $\mathbf{g}_r \in \mathbb{R}^{n_r}$, con las cuales se imponen n_r condiciones de frontera en $x = 1$.

Para encontrar estas condiciones se desacopla el sistema (B.1a) en su forma característica. En el apéndice A se vio que la variable característica \tilde{u}_k correspondiente al k -ésimo valor propio satisface

$$\frac{\partial \tilde{u}_k}{\partial t} = 0, \quad \frac{\partial x}{\partial t} = \lambda_k. \quad (\text{B.2})$$

Esto implica que la solución es constante, $\tilde{u}_k(x, t) = c_k$, a lo largo de la k -ésima característica $\mathcal{C}_k : x - \lambda_k t = c$ con c y $c_k \in \mathbb{R}$.

La figura (B.1) muestra las curvas características cuando los valores propios son positivos (izquierda) o negativos (derecha). Cuando $\lambda_k > 0$, y al tomar un punto sobre la frontera $x = 1$, vemos que la característica que pasa por dicho punto surge desde el eje x o desde el eje t . Es decir, de las condiciones iniciales o de las condiciones de frontera impuestas en $x = 0$ respectivamente. Así, únicamente es necesario dar condiciones para \tilde{u}_k en $t = 0$ y en $x = 0$. De manera similar, para el caso $\lambda_k < 0$ se deben establecer condiciones para \tilde{u}_k en $t = 0$ y $x = 1$. En conclusión, las condiciones de frontera e iniciales de la variable característica \tilde{u}_k deben colocarse en las fronteras en las cuales las curvas características inicien.

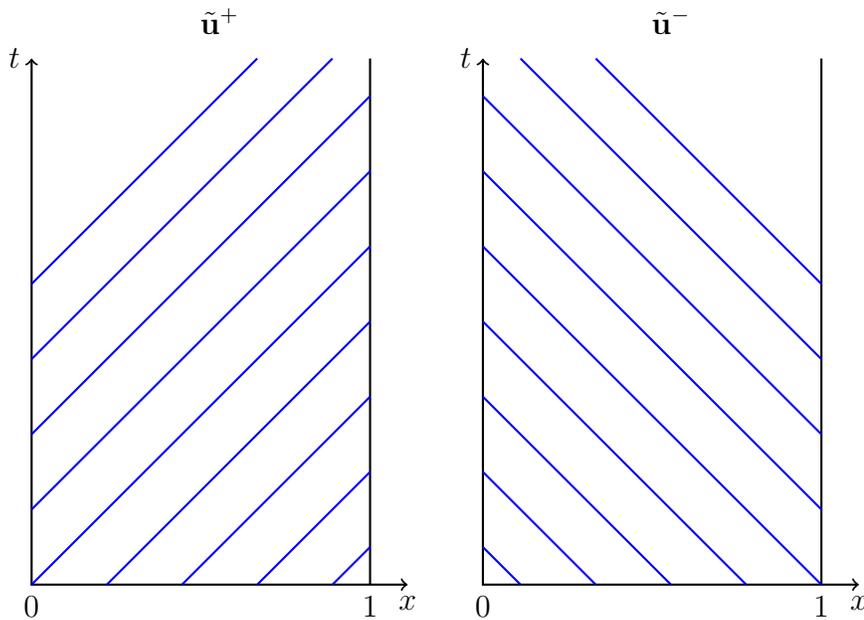


Figura B.1: Características del sistema hiperbólico, correspondientes a los casos $\lambda_k > 0$ (izquierda) y $\lambda_k < 0$ (derecha).

Ahora se reescribe el vector de variables características $\tilde{\mathbf{u}}$ como

$$\tilde{\mathbf{u}} = \begin{pmatrix} \tilde{\mathbf{u}}^- \\ \tilde{\mathbf{u}}^+ \end{pmatrix} \quad \text{donde} \quad \tilde{\mathbf{u}}^- := \begin{pmatrix} \tilde{u}_1 \\ \vdots \\ \tilde{u}_{n-p} \end{pmatrix}, \quad \tilde{\mathbf{u}}^+ := \begin{pmatrix} \tilde{u}_{n-p+1} \\ \vdots \\ \tilde{u}_n \end{pmatrix} \quad (\text{B.3})$$

Los vectores $\tilde{\mathbf{u}}^-$ y $\tilde{\mathbf{u}}^+$ son los correspondientes a los valores propios negativos y positivos, respectivamente. Además, la matriz de vectores propios derechos \mathbf{S} se puede separar como

sigue:

$$\mathbf{S} = (\mathbf{S}^- | \mathbf{S}^+) \quad \text{con} \quad \mathbf{S}^- = (\mathbf{s}_1, \dots, \mathbf{s}_{n-p}), \quad \mathbf{S}^+ = (\mathbf{s}_{n-p+1}, \dots, \mathbf{s}_n) \quad (\text{B.4})$$

donde las matrices \mathbf{S}^- y \mathbf{S}^+ contienen los vectores propios correspondientes a los valores propios negativos y positivos, respectivamente.

Con esto es posible separar las condiciones de frontera y llevarlas a las variables características, debido a que

$$\mathbf{u}(x, t) = \mathbf{S}\tilde{\mathbf{u}}(x, t).$$

Considérese primero la condición en $x = 0$, entonces:

$$\begin{aligned} \mathbf{C}_l \mathbf{u}(0, t) &= \mathbf{g}_l(t) \\ \Rightarrow \mathbf{C}_l \mathbf{S}\tilde{\mathbf{u}}(0, t) &= \mathbf{g}_l(t) \\ \Rightarrow \mathbf{C}_l (\mathbf{S}^- \tilde{\mathbf{u}}^-(0, t) + \mathbf{S}^+ \tilde{\mathbf{u}}^+(0, t)) &= \mathbf{g}_l(t) \\ \Rightarrow \mathbf{C}_l \mathbf{S}^- \tilde{\mathbf{u}}^-(0, t) + \mathbf{C}_l \mathbf{S}^+ \tilde{\mathbf{u}}^+(0, t) &= \mathbf{g}_l(t) \end{aligned}$$

De lo que se discutió antes, el término $\tilde{\mathbf{u}}^-(0, t)$ no es necesario, por lo cuál, únicamente el término correspondiente a los valores propios positivos $\tilde{\mathbf{u}}^+(0, t)$ es requerido. Para hallarlo, la matriz $\mathbf{C}_l \mathbf{S}^+$ de tamaño $n_l \times p$ debe ser invertible, esto implica que $n_l = p$. De modo que el número de condiciones de frontera en $x = 0$ debe ser igual al número de valores propios positivos. Análogamente, para la frontera $x = 1$ se llega a que la matriz $\mathbf{C}_r \mathbf{S}^-$ de tamaño $n_r \times n - p$ debe ser invertible, así $n_r = n - p$. Entonces, el número de condiciones de frontera en $x = 1$ debe ser igual al número de valores propios negativos.

En resumen, las condiciones de frontera para las variables características dependen del número de valores propios positivos y negativos y de la frontera con la que se esté tratando.

B.2. Condiciones de frontera: Caso no lineal

Los resultados de la sección anterior pueden extenderse al caso no lineal:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u})}{\partial x} = \mathbf{0}, \quad x \in (0, 1), \quad t > 0, \quad (\text{B.5a})$$

$$\mathbf{u}(x, 0) = \mathbf{v}(x), \quad x \in (0, 1), \quad (\text{B.5b})$$

$$\mathbf{C}_l(\mathbf{u})(0, t) = \mathbf{g}_l(t), \quad \mathbf{C}_r(\mathbf{u})(1, t) = \mathbf{g}_r(t), \quad t > 0, \quad (\text{B.5c})$$

tomando en cuenta algunas consideraciones más. Una de ellas es el hecho de que ahora los valores propios dependen de la solución \mathbf{u} , por lo que el signo de ellos dependerá también de \mathbf{u} . Como consecuencia, ahora el número de valores propios positivos en las fronteras se escribe como $p(\xi, t)$, con $\xi = 0, 1$. Además, \mathbf{C}_l es una función vectorial que transforma el

espacio n -dimensional en el espacio n_l -dimensional y $\mathbf{g}_l(t)$ es un vector n_l -dimensional. Del mismo modo para \mathbf{C}_r y $\mathbf{g}_r(t)$.

Otra de estas consideraciones es que ahora se utiliza la matriz de vectores propios generalizados izquierdos, $\mathbf{T} = \mathbf{S}^{-1}$. De acuerdo a lo expuesto en el apéndice A.

Para encontrar las condiciones de frontera adecuadas nuevamente la ecuación se escribe en términos de las variables características, para ello se usan las ecuaciones de Pfaff:

$$\mathbf{T}d\mathbf{u} = d\tilde{\mathbf{u}}.$$

Suponiendo que esta ecuación es integrable, entonces $\tilde{\mathbf{u}}$ puede expresarse como función de \mathbf{u} e invirtiendo esta función se obtiene:

$$\mathbf{u} = \mathcal{C}(\tilde{\mathbf{u}}) \quad (\text{B.6})$$

Al tratarse de una ecuación homogénea las variables características \tilde{u}_k son constantes a lo largo de la k -ésima curva característica \mathcal{C}_k . Del mismo modo que en el caso lineal, se llega a que las condiciones de frontera se imponen en donde las características inicien. Esto es, $\tilde{u}_k(0, t)$ y $\tilde{u}_k(1, t)$ deben darse cuando $\lambda_k(\mathbf{u})(0, t) > 0$ y $\lambda_k(\mathbf{u})(1, t) < 0$ respectivamente. Esto es equivalente a que $\tilde{\mathbf{u}}^+(0, t)$ y $\tilde{\mathbf{u}}^-(1, t)$ deben darse.

Al sustituir la relación (B.6) en las condiciones de frontera (B.5c), las condiciones de frontera en $x = 0$ se escriben como:

$$(\mathbf{C}_l \circ \mathcal{C})(\tilde{\mathbf{u}}^-, \tilde{\mathbf{u}}^+)(0, t) - \mathbf{g}_l(t) = \mathbf{0} \quad (\text{B.7})$$

donde se distinguen las variables $\tilde{\mathbf{u}}^-$ y $\tilde{\mathbf{u}}^+$. Así, se obtiene una función implícita para estas últimas y $\tilde{\mathbf{u}}^+(0, t)$ debe despejarse. De este modo, para utilizar el teorema de la función implícita, se debe pedir que la matriz

$$\frac{\partial}{\partial \tilde{\mathbf{u}}^+} (\mathbf{C}_l \circ \mathcal{C})(0, t)$$

de tamaño $n_l \times p(0, t)$ sea invertible, con $p(0, t)$ el número de valores propios positivos en $x = 0$ al tiempo t . Implicando que $p(0, t) = n_l$, que nuevamente significa que el número de condiciones en la frontera $x = 0$ es igual al número de valores propios positivos en esa frontera. Igualmente, para la frontera $x = 1$ se pide que

$$(\mathbf{C}_r \circ \mathcal{C})(\tilde{\mathbf{u}}^-, \tilde{\mathbf{u}}^+)(1, t) - \mathbf{g}_r(t) = \mathbf{0} \quad (\text{B.8})$$

con la matriz

$$\frac{\partial}{\partial \tilde{\mathbf{u}}^-} (\mathbf{C}_r \circ \mathcal{C})(1, t)$$

de tamaño $n_r \times (n - p(0, t))$ invertible. Luego $n_r = n - p(1, t)$.

B.3. Condiciones de frontera: Ecuaciones de aguas someras

Como ejemplo, ahora se aplicará lo anterior para establecer las condiciones de frontera a las ecuaciones de aguas someras, restringidas al intervalo $[0, 1]$. En el capítulo 2 se han calculado los valores propios de estas ecuaciones

$$\lambda_1 = u - c, \quad \lambda_2 = u + c, \quad u = \frac{q}{h} \quad c = \sqrt{gh},$$

los vectores propios

$$\mathbf{s}_1(\mathbf{u}) = \begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}, \quad \mathbf{s}_2(\mathbf{u}) = \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix},$$

y las variables características o invariantes de Riemann

$$\begin{aligned} w_1 &= u + 2c \\ w_2 &= u - 2c, \end{aligned}$$

con $c = \sqrt{gh}$. De este sistema se puede llegar a que

$$\begin{pmatrix} u \\ c \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(w_1 + w_2) \\ \frac{1}{4}(w_2 - w_1) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

Sustituyendo el valor de u y c en las relaciones anteriores y despejando h y q se obtiene

$$\mathbf{u} = \begin{pmatrix} h \\ q \end{pmatrix} = \frac{1}{16g}(w_2 - w_1)^2 \begin{pmatrix} 1 \\ \frac{1}{2}(w_1 + w_2) \end{pmatrix} = \mathcal{C}(\mathbf{w}). \quad (\text{B.9})$$

Derivando respecto a $\tilde{\mathbf{u}}$ se encuentra que

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}} \mathcal{C}(\mathbf{w}) &= \frac{1}{8g}(w_2 - w_1) \begin{bmatrix} -1 & 1 \\ -\frac{1}{4}(3w_1 + w_2) & \frac{1}{4}(w_1 + 3w_2) \end{bmatrix} \\ &= \frac{c}{2g} \begin{bmatrix} -1 & 1 \\ -(u - c) & u + c \end{bmatrix} \end{aligned}$$

Antes de determinar la cantidad de condiciones de frontera que se deben imponer, se debe

recordar que los signos de los valores propios dependen de la solución en el punto (x, t) , implicando que los signos pueden cambiar conforme vaya evolucionando la solución. Además, el número de condiciones en una frontera no determina la cantidad en la otra. Cada una es independiente de la otra. A pesar de esto, las condiciones de frontera se imponen considerando los valores propios de las frontera al tiempo $t = 0$. Con esto presente, el número de condiciones de frontera en cada extremo se determina como sigue

1. **Caso supercrítico.** $c \leq u$, es decir, $0 \leq \lambda_1$

Ambos valores propios son positivos. Esto significa que hay que dar dos condiciones de frontera en $x = 0$ y ninguna en $x = 1$. De donde $\mathbf{w}^+ = \mathbf{w}$. Luego

$$\frac{\partial}{\partial \mathbf{w}^+} (\mathbf{C}_l \circ \mathcal{C}) = \frac{\partial}{\partial \mathbf{w}} (\mathbf{C}_l \circ \mathcal{C}) = \frac{\partial}{\partial \mathbf{u}} \mathbf{C}_l \frac{\partial}{\partial \mathbf{w}} \mathcal{C}$$

es invertible, ya que su determinante es distinto de cero, esto es

$$\det \left(\frac{\partial}{\partial \mathbf{u}} \mathbf{C}_l \frac{\partial}{\partial \mathbf{w}} \mathcal{C} \right) = -\frac{c^3}{2g^2} \det \left(\frac{\partial}{\partial \mathbf{u}} \mathbf{C}_l \right) \neq 0 \quad (\text{B.10})$$

2. **Caso subcrítico.** $|u| < c$ Esto implica que $\lambda_1 < 0 < \lambda_2$, entonces la característica \mathcal{C}_1 entra al dominio desde la izquierda y la característica \mathcal{C}_2 entra al dominio desde la derecha. Por lo tanto w_1 se debe especificar en $x = 1$ y w_2 en $x = 0$. En otras palabras, $\mathbf{w}^+ = w_2$ y $\mathbf{w}^- = w_1$.

Entonces, para la frontera en $x = 0$ se tiene:

$$\begin{aligned} 0 \neq \frac{\partial}{\partial \mathbf{w}^+} (\mathbf{C}_l \circ \mathcal{C}) (0, t) &= \frac{\partial}{\partial w_2} (\mathbf{C}_l \circ \mathcal{C}) (0, t) \\ &= \frac{\partial}{\partial \mathbf{u}} \mathbf{C}_l \frac{\partial}{\partial w_2} \mathcal{C} (0, t) \\ &= \left(\left(\frac{\partial \mathbf{C}_l}{\partial h}, \frac{\partial \mathbf{C}_l}{\partial q} \right) \frac{c}{2g} (1, u + c)^T \right) (0, t) \\ &= \frac{c}{2g} \left(\frac{\partial \mathbf{C}_l}{\partial h} + (u + c) \frac{\partial \mathbf{C}_l}{\partial q} \right) (0, t) \end{aligned}$$

Simplificando, se obtiene:

$$\left(\frac{\partial \mathbf{C}_l}{\partial h} + (u + c) \frac{\partial \mathbf{C}_l}{\partial q} \right) (0, t) \neq 0 \quad (\text{B.11})$$

Igualmente, para la frontera derecha, $x = 1$

$$\left(\frac{\partial \mathbf{C}_r}{\partial h} + (u - c) \frac{\partial \mathbf{C}_r}{\partial q} \right) (1, t) \neq 0 \quad (\text{B.12})$$

3. **Caso supercrítico.** $u \leq -c$ Ahora ambos valores propios son negativos, entonces $\mathbf{w}^- = \mathbf{w}$. Similarmente que en el primer caso, en la frontera $x = 1$:

$$\frac{\partial}{\partial \mathbf{w}^-} (\mathbf{C}_r \circ \mathcal{C}) = \frac{\partial}{\partial \mathbf{w}} (\mathbf{C}_r \circ \mathcal{C}) = \frac{\partial}{\partial \mathbf{u}} \mathbf{C}_r \frac{\partial}{\partial \mathbf{w}} \mathcal{C}$$

es invertible. Luego

$$-\frac{c^3}{2g^2} \det \left(\frac{\partial}{\partial \mathbf{u}} \mathbf{C}_r \right) \neq 0 \quad (\text{B.13})$$

Así se han obtenido las propiedades que deben cumplir las condiciones de frontera para las ecuaciones de aguas someras.

Bibliografía

- Steven C. Chapra y Raymond P. Canale. *Métodos numéricos para ingenieros*. McGraw-Hill,, 2003.
- Guy Chavent y Bernardo Cockburn. The local projection p^0 - p^1 -discontinuous-galerkin finite element method for scalar conservation laws. *RAIRO-Modélisation mathématique et analyse numérique*, 23(4):565–592, 1989.
- Guy Chavent y G. Salzano. A finite-element method for the 1-d water flooding problem with gravity. *Journal of Computational Physics*, 45(3):307–344, 1982.
- Bernardo Cockburn, Suchung Hou, y Chi-Wang Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV. the multidimensional case. *Mathematics of Computation*, 54(190):545–581, 1990.
- Bernardo Cockburn, San-Yih Lin, y Chi-Wang Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III: one-dimensional systems. *Journal of Computational Physics*, 84(1):90–113, 1989.
- Bernardo Cockburn y Chi-Wang Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II. general framework. *Mathematics of computation*, 52(186):411–435, 1989.
- Bernardo Cockburn y Chi-Wang Shu. The Runge-Kutta local projection p^1 -discontinuous-galerkin finite element method for scalar conservation laws. *RAIRO-Modélisation mathématique et analyse numérique*, 25(3):337–361, 1991.
- Bernardo Cockburn y Chi-Wang Shu. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, págs. 173–261, 2001.
- Departamento de Matemáticas de la Universidad de Extremadura. *Apuntes de ecuaciones diferenciales*. Documento en línea, (Última vez consultado el 11/05/2017), 2013. URL <http://matematicas.unex.es/~ricarfr/EDP/LibroEDLat.pdf>.
- Vít Dolejší y Miloslav Feistauer. *Discontinuous Galerkin Method*. Springer, 2015.
- Michel Fortin y André Fortin. A new approach for the fem simulation of viscoelastic flows. *Journal of non-newtonian fluid mechanics*, 32(3):295–310, 1989.

- Luigi Fraccarollo y Eleuterio F. Toro. Experimental and numerical assessment of the shallow water model for two-dimensional dam-break type problems. *Journal of hydraulic research*, 33(6):843–864, 1995.
- Araz M. Gharangik y M. Hanif Chaudhry. Numerical simulation of hydraulic jump. *Journal of Hydraulic Engineering*, 1991. URL [http://gen.lib.rus.ec/scimag/index.php?s=10.1061/\(ASCE\)0733-9429\(1991\)117:9\(1195\)](http://gen.lib.rus.ec/scimag/index.php?s=10.1061/(ASCE)0733-9429(1991)117:9(1195)).
- Edwige Godlewski y Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Springer, 1996.
- Amiram Harten, Peter D. Lax, y Bram Van Leer. On upstream differencing and godunov-type scheme for hyperbolic conservation laws. *SIAM review*, 25(1):35–61, 1983.
- Jan S. Hesthaven y Tim Warburton. *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. Springer Science & Business Media, 2007.
- Pierre Jamet. Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain. *SIAM Journal on Numerical Analysis*, 15(5):912–928, 1978.
- Claes Johnson y Juhani Pitkäranta. An analysis of the discontinuous galerkin method for a scalar hyperbolic equation. *Mathematics of computation*, 46(173):1–26, 1986.
- Abdul A. Khan y Wencong Lai. *Modeling shallow water flows using the discontinuous Galerkin method*. CRC Press, 2014.
- Pierre LeSaint y Pierre-Arnaud Raviart. On a finite element method for solving the neutron transport equation. *Mathematical aspects of finite elements in partial differential equations*, (33):89–123, 1974.
- Randall J. LeVeque. *Numerical methods for conservation laws*. Springer Science & Business Media, 1992.
- Ben Q. Li. *Discontinuous finite elements in fluid dynamics and heat transfer*. Springer Science & Business Media, 2005.
- Robert M. M. Mattheij, Sjoerd W. Rienstra, y Jan H. M. Ten Thije Boonkkamp. *Partial differential equations: modeling, analysis, computation*. Siam, 2005.
- William H. Reed y Thomas R. Hill. Triangular mesh methods for the neutron transport equation. 1973. *Los Alamos Scientific Laboratory, Los Alamos, NM*, 1973.
- Armin Schoklitsch. Ueber dammbruchwellen. *Sitzungsberichte der Kaiserlichen Akademie Wissenschaften*, 126, 1917.
- Chi-Wang Shu. Total-variation-diminishing time discretizations. *SIAM Journal on Scientific and Statistical Computing*, 9(6):1073–1084, 1988.

- Chi-Wang Shu y Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, 1988.
- Chi-Wang Shu y Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes, ii. *Journal of Computational Physics*, 83(1):32–78, 1989.
- William Carlisle Thacker. Some exact solutions to the nonlinear shallow-water wave equations. *Journal of Fluid Mechanics*, 107:499–508, 1981.
- Eleuterio F. Toro. The dry-bed problem in shallow-water flows. Inf. téc., Cranfield Institute of Technology, 1990.
- Eleuterio F. Toro. *Shock-capturing methods for free-surface shallow flows*. John Wiley, 2001.
- Landon C. Wellford y J. Tinsley Oden. Discontinuous finite-element approximations for the analysis of shock waves in nonlinearly elastic materials. *Journal of Computational Physics*, 19(2):179–210, 1975.
- X. Ying, Abdul A. Khan, y Sam S. Y. Wang. An upwind method for one-dimensional dam break flows. En *Proceeding of XXX IAHR Congress*, págs. 245–252. 2003.