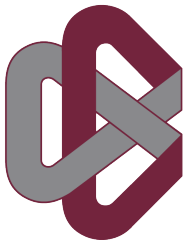


CIMAT

Estimación de la Batimetría en las Ecuaciones de Saint-Venant por el Método del Sistema Adjunto y Aproximación con el Método Galerkin Discontinuo

T E S I S

Que para obtener el grado de

Doctor en Ciencias

con Orientación en

Matemáticas Aplicadas

presenta

José Alejandro Butanda Mejía

Director de Tesis:

Dr. Miguel Ángel Moreles Vázquez

Centro de Investigación en Matemáticas, A. C.

Guanajuato, Guanajuato

08 de Abril de 2021

A mis Padres
Victor y Josefina

A mis Hermanos
Arturo, Juan y Josué

A mi Compañero de Vida
Migue

In solving a problem of this sort, the grand thing is to be able to reason backwards. That is a very useful accomplishment, and a very easy one, but people do not practise it much. In the every-day affairs of life it is more useful to reason forwards, and so the other comes to be neglected. There are fifty who can reason synthetically for one who can reason analytically. Let me see if I can make it clearer. Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds, and argue from them that something will come to pass. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backwards, or analytically.

Arthur Conan Doyle, *A Study in Scarlet*

Agradecimientos

Sé que me he olvidado de ti en estos años, pero tu siempre has guiado mi camino, me has dado fuerza para seguir adelante y has cuidado de mi y de los que amo. Gracias Dios por los grandes y excepcionales regalos que me has dado en mi vida.

Por su amor incondicional, gracias a mi familia: Victor y Josefina que han sabido ser unos padres excepcionales y me han aceptado con mis virtudes y defectos; a mis hermanos Arturo, Juan y Josué que me han dado fuerzas para seguir adelante; a Migue, mi compañero de vida, que a pesar de mis imperfecciones y errores me alenta a ser una mejor persona. Gracias por apoyarme en todos los sentidos, ya que sin ustedes no hubiera sido esto posible.

Al Centro de Investigación en Matemáticas, A. C. por haberme brindado la oportunidad de realizar mis estudios de doctorado.

Al Dr. Miguel Ángel Moreles Vázquez, gracias por su apoyo incondicional, su inmensa paciencia y entusiasmo para la culminación de este trabajo.

A mis sinodales, los doctores Salvador Botello (CIMAT), Rafael Herrera (CIMAT), Gerardo Hernández (UNAM) y Pedro González (UNAM) por sus valiosas observaciones, correcciones y sugerencias para enriquecer este trabajo.

A mis amigos: a Rafael y Oscar por todos los buenos momentos que pasamos en Guanajuato; a Hugo y Rubén, gracias a ambos por esas pláticas acompañadas de una buena cerveza.

Al Consejo Nacional de Ciencia y Tecnología por brindarme una beca para mis estudios de doctorado con número (CVU/ becario) 508129/287179.

Índice general

Introducción

En un contexto general, este trabajo de tesis se enmarca en el estudio de flujos de superficie libre, los cuales aparecen en múltiples aplicaciones, [37].

Estos flujos están modelados por las ecuaciones de Navier-Stokes. Cuando los flujos son dominados por los cambios horizontales, esto es, la componente vertical puede descartarse, se obtienen las ecuaciones de aguas someras. Si el flujo es primordialmente unidimensional, el modelo se conoce también como las ecuaciones de Saint-Venant.

El flujo de aguas someras se plantea como un problema con valores iniciales y de frontera. El problema directo consiste en establecer resultados de existencia, unicidad y estabilidad de la solución, esto es, determinar cuando es un problema bien planteado en el sentido de Hadamard. Desde la perspectiva del Análisis Numérico, el propósito es proponer y analizar métodos numéricos de aproximación. Ambos son problemas de actualidad y con un avance significativo. En el caso de métodos numéricos para encontrar una solución aproximada, un compendio se presenta en [43].

En el estudio del problema directo de las ecuaciones de aguas someras, se supone conocida toda la información: parámetros, dominio de definición, condiciones iniciales y de frontera. Esto no sucede en la práctica, solo se tiene información parcial de la solución, e.g., mediciones de componentes de velocidad horizontales en un número discreto de puntos en tiempo y espacio.

En parte, esta restricción práctica ha dado lugar al estudio de los Problemas Inversos en las diversas áreas del conocimiento. Los problemas inversos son en general mal planteados en el sentido de Hadamard, y han motivado el desarrollo de nuevas metodologías en diversas áreas de la Matemática, en particular en el Análisis Funcional No Lineal, las Ecuaciones Diferenciales, la Optimización y el Control, entre otras. También es bien conocido que la aproximación de los problemas inversos lleva a la solución numérica de problemas que son intrínsecamente mal condicionados. De la misma manera, se han pro-

puesto nuevos métodos numéricos de solución para abordar esta problemática.

En relación a las ecuaciones de aguas someras, algunos problemas inversos de interés son: dadas mediciones de la velocidad en la superficie, identificar condiciones de flujo (ondas de choque, Tirupathi et al. [39]), propiedades del fluido, batimetría, etc. Una revisión reciente del estado del arte para problemas inversos en flujos de superficie libre, y en particular de aguas someras, se presenta en [13].

Además, en las ecuaciones de aguas someras se supone el conocimiento de ciertas variables y valores como la batimetría de la región de estudio, las condiciones iniciales, las condiciones de frontera y los parámetros del modelo, las cuales no siempre son conocidas. Por ejemplo, el coeficiente de Manning describe la rugosidad del fondo de la región y generalmente se le asigna un valor constante en tablas establecidas, sin embargo este depende de diversos factores como la irregularidad del canal, la vegetación y la erosión, entre otros; es decir, varía con la posición y, en escenarios con tiempos largos, también puede cambiar. Otra variable importante es la batimetría (relieve de la superficie del fondo), que depende de la posición y el tiempo en muchas regiones, como ríos, lagos y en terrenos inundados, y es imposible de generar por medio de mediciones directas. También es de interés identificar condiciones de flujo (ondas de choque, este es tratado por Tirupathi en [39]), propiedades del fluido, etc.

En consecuencia, es necesario generar estimaciones de estas variables, utilizando información existente y datos recopilados en campo. De manera especifica, en este trabajo nos enfocamos en la reconstrucción de la batimetría asociada al problema de flujos superficiales gobernados por las ecuaciones de aguas someras en canales con paredes verticales y anchura uniforme.

Esta área de investigación es muy activa. En seguida mencionaremos contribuciones recientes que han sido útiles en el desarrollo del presente trabajo, tales como la investigación en problemas restringidos de ecuaciones diferenciales parciales hiperbólicas, prestando especial atención al sistema de ecuaciones someras en canales. Una estimación de los parámetros inherentes al problema para sistemas hiperbólicos en una dimensión basados en el método adjunto fue propuesto en [29]. Aplicaciones a estimaciones de parámetros de sistemas hidrológicos y flujos superficiales puede ser encontrado en [31] y [30], respectivamente. El problema de determinar de manera óptima las condiciones iniciales en las ecuaciones de aguas someras es tratado en [20], dando condiciones suficientes para la convergencia hacia la condición inicial verdadera. Más aún, un marco general para tratar problemas de optimización con ecuaciones diferenciales parciales hiperbólicas es presentado en [28], con aplicaciones a sistemas acoplados en un problema de ecuaciones

diferenciales parciales restringidas, la formulación del método adjunto fue presentado, y se establecen condiciones para garantizar la existencia de una solución óptima. Un enfoque numérico para reconstruir la topografía de un rio a partir de mediciones de la superficie libre es presentada en [13]. En [14], los autores consideran mediciones de velocidad y suponen un flujo constante. Recuperación de imágenes de la batimetría usando promedios de profundidad con observaciones de velocidad cuasi-estables es llevado a cabo en [23].

Un problema no abordado en la literatura, es la determinación de la batimetría usando mediciones de velocidad, recientemente se han diseñado artefactos para ello como en [5].

En consecuencia, en este trabajo proponemos una solución a este problema de recuperación de batimetría dadas mediciones de velocidad. Para ilustrar la eficacia de la metodología consideramos dos ejemplos no abordados en la literatura. Mostramos la recuperación de una batimetría ondular asociada al problema de mareas oceánicas. El problema directo asociado se presenta en [4]. Otro problema no trivial, es la recuperación de batimetrias discontinuas. En este caso reconstruimos la batimetría estudiada en varios trabajos, pero elegimos la presentación de Haegyun Lee en [22].

La metodología para la solución del problema inverso de interés es clásica. El problema se formula como un problema de minimización con restricciones. El funcional a minimizar depende de la variable desconocida y la variable de estado. El dominio de definición es un espacio normado, y la restricción sobre la variable de estado es que sea solución de las ecuaciones de Saint-Venant.

Esto nos lleva a formular un problema generalizado a espacios de Hilbert de multiplicadores de Lagrange. Usando técnicas del Análisis Funcional No Lineal desarrollamos métodos de descenso. La existencia de direcciones de descenso de un funcional diferenciable en el sentido de Fréchet, es determinando el gradiente por medio del sistema adjunto y la aplicación del Teorema de Representación de Riesz. En el proceso de demostración, utilizamos la derivada en el sentido de Gâteaux y otras técnicas de cálculo variacional. Es de notar que el sistema adjunto, es un sistema hiperbólico lineal por resolver como paso intermedio al cálculo del gradiente.

Los métodos numéricos y computacionales, son parte integral del trabajo. Antes de introducir nuestra contribución, describimos algunos antecedentes. En los últimos años se han desarrollado varios esquemas numéricos para resolver las ecuaciones de aguas someras. El método de diferencias finitas y el método de volumen finito son dos de los métodos extensamente usados para problemas que involucran las ecuaciones de aguas someras y dinámica de fluidos. Wang et al. [41] se describe un esquema tipo TVD y diferencias

finitas para resolver el problema de ruptura de presas. Lin et al. [26] uso el método de volumen finito para obtener una solución numérica a las ecuaciones de aguas someras. El método de elemento finito también se ha utilizado, en parte por su ventaja para lidiar con geometrías complejas. Sin embargo, el método de elemento finito falla para modelar el término convectivo en problemas generales de dinámica de fluidos, y se requieren técnicas más complicadas para superar estas dificultades, tales como el método de elemento finito penalizado [19], elemento finito split-characteristic [45].

En años recientes el método de Galerkin ha sido desarrollado para resolver sistemas hiperbólicos de ecuaciones diferenciales parciales. Los primeros en introducir este método fueron Reed y Hill en [34] para resolver la ecuación de transporte de neutrones, una ecuación hiperbólica independiente del tiempo. Cockburn y Shu et al. [9] han desarrollado este método para leyes conservativas incorporando métodos explícitos Runge-Kutta tipo TVD para la integración del esquema temporal. La combinación de estos métodos es conocido como método *Runge-Kutta Disconituous Galerkin* (RKDG).

El método RKDG tiene ventajas sobre los métodos de volumen finito y elemento finito. Cuando usamos elementos discontinuos podemos integrar esquemas tipo *upwind* que son usados en elemento finito para tratar con problemas convectivos dominantes. Como en el caso de elemento finito, el método RKDG puede tratar fácilmente con geometrías complejas y utilizar un alto orden de aproximación espacial. Al desacoplar los elementos a través de los flujos en las fronteras no requerimos ensamblar una matriz global y podemos aplicar esquemas explícitos en el tiempo de manera local. El método RKDG, se puede adaptar fácilmente a problemas físicos que involucren shocks y discontinuidades. En el método de Galerkin discontinuo, los elementos son desacoplados y la precisión con la que calcularemos el flujo en las fronteras determina el rendimiento y precisión de la solución que obtengamos.

De esta manera, nuestra propuesta numérico-computacional, es en base al método RKDG. Al establecer el esquema numérico de Galerkin Discontinuo para nuestro problema, surge de manera natural un problema de valores iniciales. Adaptaremos ciertas características al método DG para asegurar que el esquema temporal es estable. En la solución del sistema adjunto se requiere la aproximación en cada elemento espacial de un término que involucra el operador adjunto del operador de observación. Nuestra aproximación es un aplicación directa del teorema de Diferenciación de Lebesgue.

En este trabajo de tesis desarrollaremos las herramientas para resolver el problema inverso de recuperar la batimetría dadas observaciones de velocidad mediante un enfoque numérico. Organizamos la tesis de la siguiente manera:

- En el capítulo 1 es una breve recopilación de los fundamentos teóricos necesarios para el desarrollo de este trabajo de tesis. En la primera parte de este capitulo esta dedicado a leyes de conservación escalares, describimos el surgimiento de leyes de conservación mediante de principios físicos; a partir de consideraciones geométricas construimos el método de características, finalmente establecemos resultados sobre la existencia de soluciones entrópicas del problema de Cauchy asociado a leyes de conservación. En la segunda parte abordamos la teoría de Análisis no Lineal en espacios de Hilbert: Funcionales y propiedades; derivadas de Gâteaux y Fréchet, así como resultados de existencia, la noción de gradiente de un funcional y la descripción del método de gradiente adjunto.

- En el capítulo 2, a partir de consideraciones físicas derivamos las ecuaciones de Saint-Venant

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ u \end{bmatrix} + \frac{\partial}{\partial x}\begin{bmatrix} hu \\ \frac{1}{2}u^2 + gh \end{bmatrix} = \begin{bmatrix} 0 \\ -gB_x \end{bmatrix}, \quad x \in [a,b], \quad t \geq 0.$$

  donde h y u son la altura y la velocidad respectivamente, $B(x)$ es la batimetría y g es la constante gravitacional; mencionamos algunas de sus propiedades y establecemos el problema de nuestro interés: Hallar el mínimo del funcional $\mathcal{J}(B) : \mathbb{L}^2[a,b] \to \mathbb{R}$ dado por

$$\mathcal{J}(B) = \frac{1}{2}\sum_{j,n}\left[u(x_j, t_n; B) - \hat{u}_{j,n}\right]^2,$$

  donde $\hat{u}$ es un conjunto de datos observados tomados en una malla $\{(x_j, t_n)\} \subset [a,b]$, $j = 0, 1, \ldots, N$, $n = 0, 1, \ldots, M$.

- En el capítulo 3 describimos de manera sucinta el método de optimización de descenso continuo que será imprescindible en establecer un esquema numérico para resolver nuestro problema. En las últimas secciones de este apartado, describimos dos aportaciones significativas: Una expresión analítica para $\nabla\mathcal{J}(B)$ y una aproximación numérica para el operador adjunto del operador de observación.

- En el capítulo 4 describimos el método numérico de Galerkin Discontinuo. En una primera instancia desarrollamos este método para leyes de conservación escalar; describimos el esquema temporal tipo TVD Runge-Kutta y realizamos pruebas numéricas conociendo a priori la solución analítica de un problema estándar, y de esta manera calculamos errores y razón de convergencia usando la norma en $\mathbb{L}^2$. En la segunda parte del capítulo aplicamos el método de Galerkin Discontinuo para el sistema de ecuaciones de Saint-Venant y describimos un algoritmo con pseudo código para encontrar una solución numérica. En la última sección describimos el método

de Galerkin Discontinuo para el problema adjunto con el fin de hallar $\nabla \mathcal{J}(B)$, en este punto son fundamentales los resultados del capítulo 3.

- En el capítulo 5 mostramos los resultados numéricos que hemos obtenido para el problema directo y adjunto para cuatro problemas:

  Solución Analítica En el caso de que tengamos una batimetría nula se conocen ciertas soluciones analíticas, este ejemplo lo usamos para observar la convergencia del método de Galerkin discontinuo. Esta situación es tratada en [17].

  Tope Para este ejemplo hemos considerado una topografía en forma de tope. En el caso del problema directo, esta ampliamente documentado en la literatura.

  Mareas Oceánicas En este modelo describimos el comportamiento de mareas en aguas poco profundas, el problema directo es tratado en [4].

  Batimetría Discontinua Este ejemplo es una modificación del anterior con una batimetría discontinua, el problema directo y un método numérico son descritos en [22].

- En el capítulo 6 planteamos el problema inverso para las ecuaciones de aguas someras bidimensionales, exponemos el método adjunto para esta situación y establecemos una expresión analítica para el gradiente.

Capítulo 1

Preliminares de Leyes de Conservación y Análisis no Lineal

En este capitulo abordaremos los conceptos teóricos tomados en cuenta para el desarrollo del presente trabajo. Estos tópicos se abordan profundamente en [8], [10] y [27].

1.1. Leyes de Conservación Escalares

Una ley de conservación escalar escrita en su forma conservativa es una ecuación diferencial parcial de primer orden homogénea de la forma

$$\partial_t u + \partial_x f(u) = 0. \tag{1.1}$$

Con el fin de entender físicamente el significado de la ecuación anterior, consideremos una solución suave a (1.1).

Al usar la fórmula de Green en un intervalo $[a, b]$, obtenemos

$$\begin{aligned} \frac{d}{dt}\left(\int_a^b u(x,t)dx\right) &= \int_a^b \partial_t u(x,t)dx \\ &= -\int_a^b \partial_x f(u(x,t))dx \\ &= f(u(a,t)) - f(u(b,t)). \end{aligned} \tag{1.2}$$

Como consecuencia, la cantidad medida por $u(x,t)$ contenida en $[a,b]$, esto es $\int_a^b u(x,t)dx$, solo puede cambiar debido al flujo $f(u(a,t))$ de u a través del punto $x = a$ y al flujo $f(u(b,t))$ de u a través del punto $x = b$. En otras palabras, la cantidad u no se crea o se destruye. Por esta razón, es natural referirse a u como una cantidad conservada y a f como una función de flujo relacionada a la ecuación (1.1).

En particular, cuando $\lim_{|x|\to\infty} u(x,t) = 0$ para todo $t \in \mathbb{R}^+$ y el flujo es normalizado

de tal manera que $f(0) = 0$, podemos hacer tender a a $-\infty$ y b a ∞ en (1.2) y obtener que la integral de la cantidad conservada sobre todo el espacio es independiente del tiempo.

De manera más concreta ilustramos como surgen las leyes de conservación a partir de principios físicos, consideremos el caso fundamental para derivar relaciones de conservación de masa en una dimensión. Para tal propósito estudiaremos el flujo de un gas a través de un tubo.

Sea x la distancia a lo largo de un tubo y sea $\rho(x,t)$ la densidad del gas en el punto x al tiempo t. Esta densidad es definida de tal manera que la masa total del gas en cualquier sección entre a y b, está dada por la integral:

$$\text{masa en } [a,b] \text{ al tiempo } t \ = \int_a^b \rho(x,t)dx.$$

Si suponemos que las paredes del tubo son impermeables y la masa no es creada ni destruida, entonces la masa en una sección de referencia solo puede cambiar debido al flujo del gas a lo largo de los puntos extremos a o b.

Sea $u(x,t)$ la velocidad del gas en el punto x al tiempo t, entonces la razón del flujo de gas, a través de este punto está dado por

$$\text{flujo en } (x,t) = \rho(x,t)u(x,t).$$

Por los comentarios anteriores, la razón de cambio de masa en $[a,b]$ es dada por la diferencia de los flujos en a y b, esto es,

$$\frac{d}{dt}\int_a^b \rho(x,t)dx = \rho(b,t)u(b,t) - \rho(a,t)u(a,t). \tag{1.3}$$

La integral anterior establece una ley de conservación. Otra forma es obtenida integrando (1.3) en el intervalo temporal $[t_1,t_2]$, dando una expresión para la masa en $[a,b]$ al tiempo t_2 en términos de la masa al tiempo t_1 y el flujo en cada frontera durante este lapso, esto es,

$$\int_a^b \rho(x,t_2)dx = \int_a^b \rho(x,t_1)dx + \int_{t_1}^{t_2} \rho(a,t)u(a,t)dt - \int_{t_1}^{t_2} \rho(b,t)u(b,t)dt. \tag{1.4}$$

Para derivar la forma diferencial de la ley de conservación, debemos suponer que $\rho(x,t)$ y $u(x,t)$ son funciones diferenciables. Entonces, usando

$$\rho(x,t_2) - \rho(x,t_1) = \int_{t_1}^{t_2} \frac{\partial}{\partial t}\rho(x,t)dt$$

y

$$\rho(b,t)u(b,t) - \rho(a,t)u(a,t) = \int_a^b \frac{\partial}{\partial x}\rho(x,t)u(x,t)dx$$

en (1.4) obtenemos

$$\int_{t_1}^{t_2}\int_a^b \left[\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\rho(x,t)u(x,t)\right] dxdt = 0. \tag{1.5}$$

Debido a que esta relación se verifica para cualquier sección $[a,b]$ y sobre cualquier intervalo de tiempo $[t_1,t_2]$, concluimos que el integrando en (1.5) debe ser idénticamente cero, es decir,

$$\rho_t + (\rho u)_x = 0. \tag{1.6}$$

Esta es la forma diferencial para la ley de conservación de masa. Es posible calcular una solución a la ley de conservación (1.6) solo si la velocidad $u(x,t)$ es conocida a priori o es una función de $\rho(x,t)$. Si este es el caso, entonces ρu es de al forma $\rho u = f(\rho)$, y la ecuación (1.6) se convierte es una ley de conservación escalar para ρ,

$$\rho_t + f(\rho)_x = 0. \tag{1.7}$$

De manera más frecuente la ecuación (1.6) se debe resolver en conjunto con una ecuación que establece la conservación de momento.

Una generalización natural de la ley de conservación (1.1) es

$$\partial_t u + \partial_x f(x,t,u) = g(x,t,u). \tag{1.8}$$

La ecuación (1.8) es referida como una ley de balance, donde g represente el término fuente.

Si $f \in \mathcal{C}^1(\mathbb{R})$ y u es una solución suave de (1.1), entonces es posible aplicar la regla de la cadena y podemos reescribir la ecuación (1.1) en su forma cuasilineal

$$\partial_t u + a(u)\partial_x u = 0, \tag{1.9}$$

donde $a = f' \in \mathcal{C}^0(\mathbb{R})$ es la derivada de f.

En el caso de que tengamos soluciones suaves, las ecuaciones (1.1) y (1.9) son equivalentes. Sin embargo, si u tiene un salto, la ecuación cuasilineal (1.9) en general no está bien definida, ya que es involucrado un producto de una función discontinua, $a(u)$, con una medida de Dirac, $\partial_x u$. Por lo tanto, (1.9) es válida solo para la clase de funciones continuas, mientras que (1.1) puede ser interpretada en el sentido de distribuciones (para

más detalles, puede consultarse [42]).

Observemos que la generalización de (1.9) es representada por la ecuación de transporte

$$\partial_t u + a(x,t,u)\partial_x u = g(x,t,u),$$

ver por ejemplo [24] para más detalles.

Para una condición inicial $u_0 : \mathbb{R} \to \mathbb{R}$ nos interesa estudiar el problema de Cauchy (problema de valores iniciales) para la ecuación (1.1), es decir,

$$\begin{aligned} \partial_t u + \partial_x f(u) &= 0, \\ u(x,0) &= u_0(x). \end{aligned} \tag{1.10}$$

Un estudio clásico de este problema se presenta en [11].

El problema de Cauchy más simple es el llamado problema de Riemann, que corresponde a una condición inicial del tipo

$$u_0(x) = \begin{cases} u_l & \text{si } x < 0 \\ u_r & \text{si } x > 0, \end{cases}, \quad u_l \neq u_r.$$

Las únicas soluciones entrópicas de este problema de Cauchy dan origen a ondas de choque o de rarefacción (dependiendo si $u_l < u_r$ o $u_r < u_l$). En [11] se estudia ampliamente este problema.

Definición 1.1.1 *Sean $f, u_0 : \mathbb{R} \to \mathbb{R}$ funciones de clase $\mathcal{C}^1(\mathbb{R})$. Entonces $u : \mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}$ es una solución global del problema de Cauchy* (1.10) *si es de clase $\mathcal{C}^1(\mathbb{R})$ y satisface* (1.10) *puntualmente.*

En la siguiente sección introduciremos el método de características, una técnica para resolver el problema de Cauchy (1.10) en el espacio de funciones suaves. Como veremos, en general no existen soluciones globales que sean suaves para dicho problema de Cauchy más allá de un intervalo finito de tiempo, inclusive cuando la condición incisal u_0 es suave.

Método de Características

En esta sección discutiremos brevemente una técnica para resolver leyes de conservación. La ventaja de este método es reducir una ecuación diferencial parcial a través de consideraciones geométricas a un sistema de ecuaciones diferenciales ordinarias con condiciones iniciales adecuadas.

Definición 1.1.2 *Sea u una solución suave al problema de Cauchy* (1.10) *y $x_0 \in \mathbb{R}$. La curva característica $[t \to x(x_0;t)]$ asociada a u y partiendo del punto $(x,t) = (x_0,0)$ es la solución del siguiente problema de Cauchy*

$$x'(x_0,t) = a\left(u(x(x_0,t),t)\right), \quad x(x_0;0) = x_0. \tag{1.11}$$

Proposición 1.1.3 *Sea u una solución suave de* (1.10) *y $x_0 \in \mathbb{R}$. Entonces u es constante a lo largo de la curva característica* (1.11)*, que es descrita con la siguiente expresión*

$$x(x_0;t) = a(u_0(x_0))t + x_0. \tag{1.12}$$

Esta importante propiedad nos da una manera de construir soluciones suaves. En efecto, si podemos invertir la relación (1.12) y escribimos $x_0 = x_0(x,t)$, entonces

$$u(x,t) = u_0(x_0(x,t)) \tag{1.13}$$

es una solución suave al problema de Cauchy (1.10). Este procedimiento es conocido como el método de características y es usado para construir soluciones regulares al problema de Cauchy de leyes de conservación escalares, aplicaremos este método en algunos ejemplos.

Hay muchos supuestos en la construcción anterior. En efecto, la solución a (1.11) puede ser local e imposible de invertir (1.12). En la siguiente proposición, planteamos condiciones suficientes para la existencia y unicidad de soluciones regulares y globales en el tiempo.

Proposición 1.1.4 *Sea $u_0 \in \mathcal{C}^1(\mathbb{R})$ y su derivada acotadas. Si $a \circ u_0 : \mathbb{R} \to \mathbb{R}$ es creciente, entonces la función* (1.13) *está bien definida para todo $(x,t) \in \mathbb{R} \times \mathbb{R}_+^*$ y es la única solución suave y global para el problema de Cauchy* (1.10).

Ejemplo 1.1.1 *En el caso $f(u) = au$ para $a \in \mathbb{R}$, entonces* (1.1) *se convierte en la ecuación de advección lineal y el problema de Cauchy correspondiente es*

$$\begin{aligned} \partial_t u + a\partial_x u &= 0, \\ u(x,0) &= u_0(x). \end{aligned} \tag{1.14}$$

En este caso, es posible invertir la relación (1.12) *y obtener $x_0(t) = x - at$. Si u_0 es de clase $\mathcal{C}^1$, entonces la onda viajera $u(x,t) = u_0(x-at)$ es una solución global y regular a* (1.14).

Perdida de Regularidad

A diferencia de problemas lineales, en el caso no lineal aparecen nuevas características tales como la ocurrencia de soluciones discontinuas. En efecto, la no linealidad implica que la velocidad de propagación de una onda no es constante. De esta manera, en general,

una solución puede experimentar una onda adelantada, que resulta en el surgimiento de discontinuidades, inclusive si la condición inicial es suave. Ilustraremos este aspecto estudiando el siguiente ejemplo (para situaciones más generales se puede consultar [36]).

Ejemplo 1.1.2 *Consideremos el problema de Cauchy para la ecuación de Burgers*

$$\begin{aligned} \partial_t u + \partial_x (u^2/2) &= 0, \\ u(x,0) &= \frac{1}{1+x^2}. \end{aligned} \tag{1.15}$$

Si u es una solución suave, (1.15) *es equivalente a la ecuación*

$$\partial_t u + u \partial_x u = 0.$$

Consideremos la curva característica $[t \to x(t)]$ en el plano (x,t) que parte del punto $(x,t) = (x_0, 0)$, es decir, una solución del problema de Cauchy

$$x'(t) = u(x(t),t), \quad x(0) = x_0. \tag{1.16}$$

Usando la proposición 1.1.3, la solución a (1.16) *es*

$$x(t) = x_0 + \frac{t}{1+x_0^2} \tag{1.17}$$

y la solución del problema de Cauchy es dada implícitamente por

$$u\left(x + \frac{t}{1+x^2}, t\right) = \frac{1}{1+x^2}. \tag{1.18}$$

En la gráfica 1.1, esbozamos las rectas características (1.17) *correspondientes a diferentes*

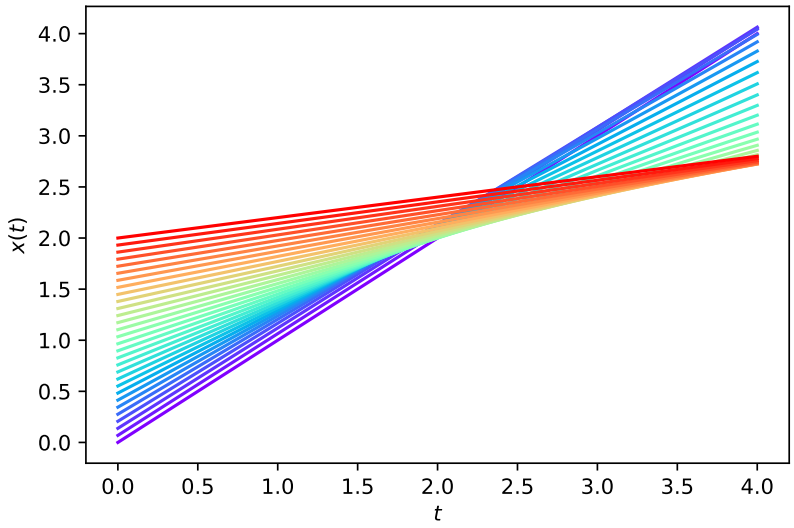


Figura 1.1: Rectas características (1.17) para distintos valores de x_0.

valores de x_0. Estas gráficas se intersecan en un tiempo $t = T$. Para calcular el valor exacto de T, es suficiente considerar las posibles intersecciones de dos características que parten de x y y, con $x \neq y$. Obtenemos que las rectas se intersecan para

$$t = \frac{(1+x^2)(1+y^2)}{x+y},$$

que representa una superficie en el plano (x, y, t). Fácilmente calculamos el (único) mínimo que ocurre en el punto $(x, y, t) = (1/\sqrt{3}, 1/\sqrt{3}, 8/\sqrt{27})$, y por tanto $T = 8/\sqrt{27}$.

Por otro lado, cuando $t > T$, las rectas características se intersecan, y el mapeo

$$x \to x + \frac{t}{1+x^2}$$

no es uno a uno y (1.18) no define una función simplemente valuada. En la gráfica 1.2 se esboza la función multivaluada u. En consecuencia, no existe una solución suave más allá del tiempo $t = T$. La única posibilidad es extender la noción de solución para todo tiempo $t \geq 0$ en el sentido de distribuciones.

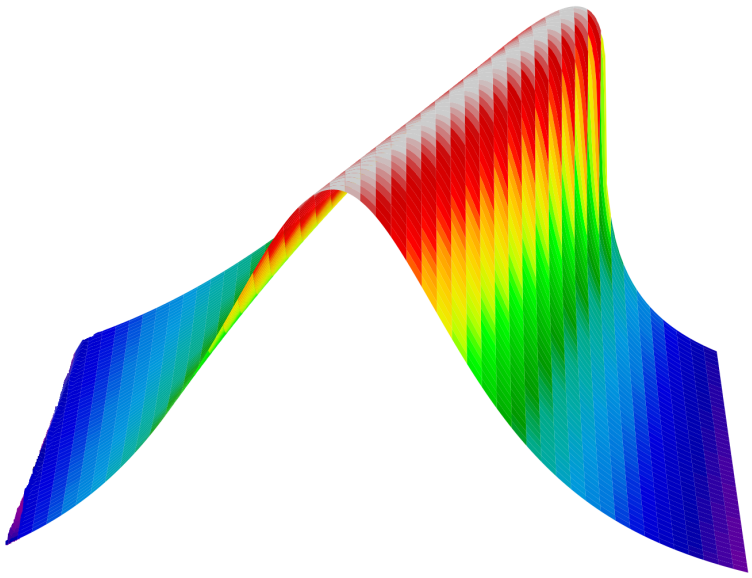


Figura 1.2: Representación de la función multivaluada u.

Si observamos la figura 1.3, los puntos sobre la gráfica de $u = u(t)$ se mueven horizontalmente con velocidad igual a la distancia al eje x. Como consecuencia, cuando t se acerca al valor critico T tenemos que

$$\lim_{t \nearrow T} \left\{ \inf_{x \in \mathbb{R}} \partial_x u(x,t) \right\} = -\infty,$$

para $t = T$ la gráfica se "pliega" y para $t > T$ hay algún x al cual hemos asociado tres

valores de u. Para extender la solución para valores de t más grandes que T tenemos que elegir alguno de estos tres valores de u. De esta manera, observamos que es imposible extender la solución y mantenerla continua.

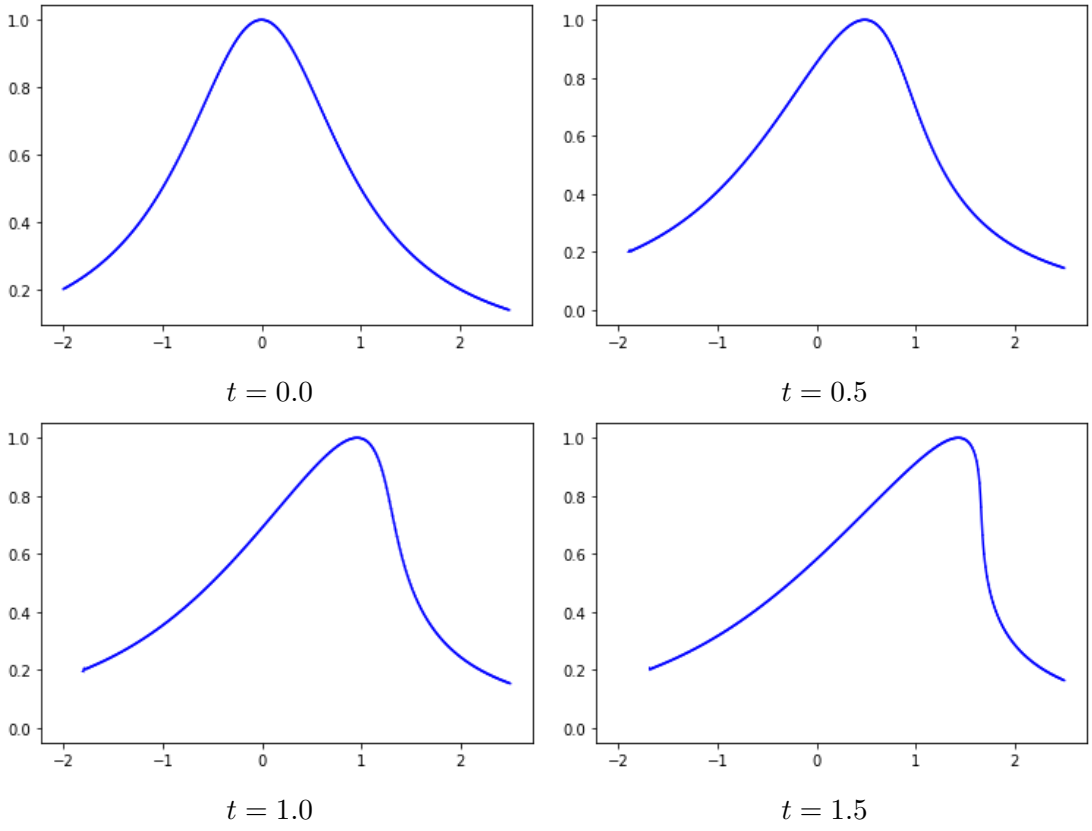


Figura 1.3: Representación en el plano (x, u) de la función multivaluada (1.18).

Proposición 1.1.5 *Sea $u_0 \in \mathcal{C}^1(\mathbb{R})$ y su derivada acotada. Entonces existe una única solución regular $u \in \mathcal{C}^1(\mathbb{R}, [0, T_{\text{máx}}))$ a (1.10) donde*

$$T_{\text{máx}} = \begin{cases} \infty & \text{si } a \circ u_0 \text{ es creciente} \\ -\left(\inf_{x \in \mathbb{R}} \left\{ \frac{d}{dx} a(u_0(x)) \right\} \right)^{-1} & \text{otro caso,} \end{cases}$$

y no existe alguna otra solución regular en un intervalo temporal más extenso.

En resumen, al usar el método de las características podemos construir una solución suave a (1.10) al menos para algún tiempo finito. Por otro lado, hemos visto que en el caso no lineal $a'(u) \neq 0$ pueden surgir discontinuidades en un tiempo finito.

Soluciones de Entropía

En esta sección estableceremos la existencia de soluciones del problema de Cauchy

$$\begin{aligned} \partial_t u + \partial_x f(u) &= 0, \\ u(x,0) &= u_0(x) \end{aligned} \tag{1.19}$$

bajo cierta suposiciones de regularidad de la condición inicial u_0 y el flujo f.

Se ha observado que el estudio clásico de leyes de conservación da lugar a problemas mal planteados. Una solución es restringir la noción de solución como sigue:

Definición 1.1.6 *Sean $u(x,t)$ y $u_0(x)$ funciones acotadas y medibles. Decimos que $u(x,t)$ es una solución débil del problema de Cauchy* (1.19) *si para toda $\phi \in \mathcal{C}_0^1(\mathbb{R})$ tenemos que*

$$\iint\limits_{t\geq 0} (u\phi_t + f(u)\phi_x)dxdt + \int\limits_{t=0} u_0\phi dx = 0.$$

Con esta noción de solución resolvemos el problema de existencia, pero no el de unicidad. Requerimos una propiedad adicional conocida como condición de entropía. A continuación enunciaremos el resultado fundamental sobre la relación de existencia y entropía.

Teorema 1.1.7 *Sea $u_0 \in \mathbb{L}^\infty(\mathbb{R})$, y $f \in \mathcal{C}^2(\mathbb{R})$ con $f'' > 0$ en $\{u : \ |u| \leq \|u_0\|_\infty\}$. Entonces existe una solución débil u de* (1.19) *con las siguientes propiedades:*

1. *$|u(x,t)| \leq \|u_0\|_\infty \equiv M$, $(x,t) \in \mathbb{R} \times \mathbb{R}^+$.*
2. *Condición de Entropía: Existe una constante $C > 0$, que depende solo de M,*

$$\mu = \min\left\{f''(u) : \ |u| \leq \|u_0\|_\infty\right\}$$

y

$$A = \max\left\{|f'(u)| : \ |u| \leq \|u_0\|_\infty\right\},$$

tal que para toda $a > 0$, $t > 0$, y $x \in \mathbb{R}$,

$$\frac{u(x+a,t) - u(x,t)}{a} < \frac{C}{t}. \tag{1.20}$$

3. *u es estable y depende continuamente de u_0 en el siguiente sentido: Si $u_0, v_0 \in \mathbb{L}^\infty(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$ con $\|v_0\|_\infty \leq \|u_0\|_\infty$, y v es solución al problema* (1.19) *con correspondiente condición inicial v_0, entonces para todas $x_1, x_2 \in \mathbb{R}$ con $x_1 < x_2$, y para cada $t > 0$,*

$$\int\limits_{x_1}^{x_2} |u(x,t) - v(x,t)|dx \leq \int\limits_{x_1 - At}^{x_2 + At} |u_0(x) - v_0(x)|dx. \tag{1.21}$$

La demostración de este teorema es constructiva y puede consultarse en [38].

A continuación mencionaremos algunas observaciones sobre este teorema. En primer lugar, de la relación (1.21) podemos inferir que la solución es estable y única, pero existe la posibilidad de la existencia de otra solución que no satisfaga (1.21). Se puede demostrar que la condición de entropía (1.20) implica unicidad. La propiedad 1 no es válida para sistemas; en efecto, la existencia de soluciones para sistemas se desprende a través estimaciones con la norma del supremo. La relación (1.20) puede ser interpretada como un teorema de regularidad, en el sentido de que implica que para cada $t > 0$, la solución $u(\cdot, t)$ es localmente de variación acotada. Para ver esto, sea $c_1 > C/t$, y consideremos $v(x,t) = u(x,t) - c_1 x$. Entonces, si $a > 0$, (1.20) implica

$$v(x+a,t) - v(x,t) = u(x+a,t) - u(x,t) - c_1 a \leq a\left(\frac{C}{t} - c_1\right) < 0,$$

por lo tanto, v es una función decreciente y de variación acotada localmente. Ya que lo mismo es cierto para la función $c_1 x$, podemos inferir que nuestra afirmación es válida para u. Así, aunque los datos iniciales pertenezcan al espacio $\mathbb{L}^\infty(\mathbb{R})$ en $t = 0$, la solución inmediatamente se vuele bastante regular para $t > 0$. De esta manera, podemos concluir que $u(\cdot, t)$ tiene un número contable de discontinuidades, y es diferenciable casi en todas partes. Esta propiedad de regularidad para la solución es, por supuesto, un fenómeno puramente no lineal.

Finalmente, observamos que la relación (1.21) demuestra que la solución tiene velocidad finita de propagación; esto se sigue al hacer $v_0 \equiv v \equiv 0$ en (1.21).

1.2. Análisis no Lineal

Método adjunto

Supongamos que $u = (u_1, \ldots, u_n)$ es la solución de n ecuaciones, no necesariamente lineales, continuas o discretas, posiblemente con valores iniciales o de frontera. Estas ecuaciones podrían incluir m variables de control $p = (p_1, \ldots, p_m)$. Con frecuencia deseamos optimizar alguna función escalar $g(u, p)$. El método adjunto nos proporcionará una manera eficiente de calcular

$$\frac{dg}{dp} = \left(\frac{dg}{dp_1}, \ldots, \frac{dg}{dp_m}\right).$$

Estas m derivadas detectan la sensibilidad de g con respecto a los cambios en p. Para una elección de p dada, dg/dp nos proporciona una dirección de búsqueda para mejorar g (es la dirección del gradiente). La dificultad a la que nos enfrentamos es que dg/dp involucra

matrices, generalmente grandes, en efecto, usando la regla de la cadena:

$$\text{Gradiente de } g(u,p): \quad \frac{dg}{dp} = \frac{\partial g}{\partial u}\frac{\partial u}{\partial p} + \frac{\partial g}{\partial p}. \tag{1.22}$$

El método adjunto nos proporciona una manera de evitar el calculo de las $n \times m$ derivadas parciales $\partial u_i/\partial p_j$. Esto reduce el esfuerzo computacional cuando p incluye miles de parámetros (m muy grande) o incluso pertenece a un espacio de dimensión infinita. El objetivo del método adjunto es calcular dg/dp resolviendo una sola ecuación adjunta en vez de m ecuaciones directas.

Sistemas Lineales

Consideremos un sistema lineal de n ecuaciones, $Au = b$, donde b depende de los parámetros $p = (p_1, \ldots, p_m)$. La función escalar $g(u,p) = c^T u$ para un vector fijo c (por ejemplo, para la elección de $c = (1, \ldots, 1)$, optimizamos $\sum u_i$). En este caso $dg/dp = 0$ y $\partial g/\partial u = c^T$. De esta manera, lo que deseamos conocer es

$$\nabla g(u,p) = c^T \frac{\partial u}{\partial p}.$$

Cada $\partial u/\partial p_j$, lo podemos obtener a partir de

$$A\frac{\partial u}{\partial p_j} = \frac{\partial b}{\partial p_j}, \tag{1.23}$$

es decir, tenemos que resolver un sistema lineal de $n \times n$ para encontrar cada uno de los m vectores $\partial u/\partial p_j$. Al sustituir (1.23) en (1.22)

$$\frac{\partial g}{\partial u}\frac{\partial u}{\partial p} = c^T \frac{\partial u}{\partial p} = c^T A^{-1}\frac{\partial b}{\partial p}.$$

La idea principal del método adjunto es calcular $c^T A^{-1}$. De esta manera, podemos resolver un solo sistema lineal en lugar de m sistemas. En efecto, sea $\lambda \in \mathbb{R}^n$ solución del sistema adjunto

$$A^T\lambda = c \quad \text{entonces} \quad \lambda^T = c^T A^{-1}. \tag{1.24}$$

Problemas no Lineales

Supongamos que tenemos n ecuaciones no lineales de la forma $f(u,p) = 0$. Estas ecuaciones determinan $u_1, \ldots, u_n$ para ciertas variables de control $p_1, \ldots, p_m$. La función escalar que consideremos, $g(u,p)$ puede ser no lineal. Una vez más, deseamos conocer el gradiente dg/dp, a fin de conocer la sensibilidad al variar p. Las derivadas de g en la ecuación (1.22) involucran a $\partial u/\partial p$. Al diferenciar $f(u,p) = 0$, obtenemos una matriz de

$n \times m$:

$$\frac{\partial f}{\partial u}\frac{\partial u}{\partial p} + \frac{\partial f}{\partial p} = 0.$$

Al sustituir en la ecuación (1.22), obtenemos una expresión para dg/dp:

$$\frac{dg}{dp} = \frac{\partial g}{\partial p} - \frac{\partial g}{\partial u}\left(\frac{\partial f}{\partial u}\right)^{-1}\left(\frac{\partial f}{\partial p}\right). \tag{1.25}$$

La mejor manera de calcular dg/dp es primero resolver el sistema lineal adjunto

$$\left(\frac{\partial f}{\partial u}\right)^{T}\lambda = \left(\frac{\partial g}{\partial u}\right)^{T}. \tag{1.26}$$

Al sustituir la solución de (1.26) en (1.25)

$$\frac{dg}{dp} = \frac{\partial g}{\partial p} - \lambda^{T}\left(\frac{\partial f}{\partial p}\right).$$

Hemos logrado evitar la multiplicación de dos matrices de dimensiones altas.

El operador adjunto de una matriz (su transpuesta conjugada) es definido como $\langle Au, \lambda\rangle = \langle u, A^T\lambda\rangle$. El adjunto de $A = d/dx$ es $A^T = -d/dx$ (usando integración por partes). Recordemos que estamos evaluando funciones escalares g, lo cual lo podemos hacer de dos maneras equivalentes como $g = c^T u$ o $g = \lambda^T b$ (manera directa o adjunta):

$$\text{Hallar } g = c^T u \quad \text{dado } Au = b \qquad \text{Hallar } g = \lambda^T b \quad \text{dado } A^T\lambda = c.$$

La equivalencia se desprende ya que $\lambda^T b = \lambda^T(Au) = (A^T\lambda)^T u = c^T u$. La elección se vuelve relevante cuando tenemos m vectores b y u, y l vectores λ y c. Para matrices grandes el esfuerzo es al resolver sistemas lineales, por lo tanto el enfoque del método adjunto es mejor cuando $l << m$.

Multiplicadores de Lagrange

Las alternativas que hemos mencionada aparecen cuando consideramos las m variables como multiplicadores de Lagrange. Las m restricciones en el problema directo son $Au = b$. En el caso no lineal son $f(u, p) = 0$. La función lagrangiana $\mathcal{L}(u, p, \lambda)$ la construimos alrededor de este tipo de restricciones:

$$\mathcal{L} = g - \lambda^T f, \qquad \frac{d\mathcal{L}}{du} = \frac{dg}{du} - \lambda^T\frac{\partial f}{\partial u}.$$

La ecuación adjunta $d\mathcal{L}/du = 0$ determina λ, como lo hemos discutido en (1.26). Cuando estudiamos la sensibilidad de $\mathcal{L}$ con respecto de p, y usando la ecuación (1.25):

$$d\mathcal{L} = \frac{d\mathcal{L}}{du}du + \frac{d\mathcal{L}}{dp}dp = \left(\frac{dg}{dp} - \lambda^T \frac{\partial f}{\partial p}\right) dp.$$

Dado un conjunto de parámetros p, estamos interesados en minimizar la función objetivo $g(u,p)$. Recordemos que la variable u está dada implícitamente por $f(u,p)$. Estas ecuaciones no lineales, y la ecuación lineal adjunta para λ, son sistemas generalmente grandes. Debemos seleccionar un algoritmo iterativo para encontrar un conjunto de parámetros p^* minimizante. Existen una gran variedad de métodos de optimización como lo son el método de descenso, método de Newton o el muy conocido método BFGS.

Cuando estudiamos problemas inversos aunados a la optimización de un conjunto de parámetros, el mayor problema al que nos enfrentamos es el alto costo computacional, ya que en cada iteración necesitamos calcular numéricamente las derivadas parciales para g de ciertos ordenes.

Diferenciabilidad Gâteaux y Fréchet

El problema que nos concierne es la minimización de un funcional, así su dominio es un espacio de dimensión infinita y requerimos del cálculo en espacios de Banach, de esta manera, en esta sección desarrollaremos de manera sucinta las nociones básicas de diferenciabilidad para funciones $f : X \to Y$ entre dos espacios de Banach X y Y.

Definición 1.2.1 *Una función f se dice Gâteaux diferenciable en x si existe un operador lineal $T_x \in \mathcal{B}(X,Y)$ tal que para todo $v \in X$,*

$$\lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} = T_x v.$$

El operador T_x es llamado la derivada de Gâteaux de f en x.

Si para algún v fijo el límite

$$\delta_v f(x) := \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

existe, decimos que f tiene derivada direccional en x en la dirección v. Por lo tanto f es Gâteaux diferenciable en x, si y solo si todas las derivadas direccionales $\delta_v f(x)$ existen y forman un operador lineal acotado $Df(x) : v \to \delta_v f(x)$.

Si el límite (en el sentido de la derivada de Gâteaux) existe uniformemente en v sobre la esfera unitaria de X, decimos que f es Fréchet diferenciable en x y T_x es la derivada

de Fréchet de f en x. Equivalentemente, si establecemos $y = tv$ entonces $t \to 0$ si y solo si $y \to 0$. Así f es Fréchet diferenciable en x si para todo y,

$$f(x+y) - f(x) - T_x(y) = o(\|y\|).$$

Si la derivada de Gâteaux existe es única, debido a que si el límite en la definición existe es único.

Una función que es Fréchet diferenciable en un punto es continua en dicho punto, pero no es el caso para funciones Gâteaux diferenciables (inclusive en el caso de dimensión finita). Por ejemplo, la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(0,0) = 0$ y $f(x,y) = x^4y/(x^6 + y^3)$ para $x^2 + y^2 > 0$ que tiene derivada de Gâteaux cero en el origen, pero no es continua en ese punto. Esta situación también nos proporciona un ejemplo de una función que es Gâteaux diferenciable pero no es Fréchet diferenciable. Otro ejemplo es el siguiente: Si X es un espacio de Banach, y $\varphi \in X^*$ es un funcional lineal discontinuo, entonces la función $f(x) = \|x\| \varphi(x)$ es Gâteaux diferenciable en $x = 0$ con derivada 0, pero $f(x)$ no es Fréchet diferenciable ya que φ no tiene límite cero en $x = 0$.

Proposición 1.2.2 (Fórmula del Valor Medio) *Si f es Gâteaux diferenciable, entonces*

$$\|f(y) - f(x)\| \leq \|x - y\| \sup_{0 \leq \theta \leq 1} \|Df(\theta x + (1-\theta)y\| \,.$$

Claramente la diferenciabilidad Fréchet tiene las propiedades de adición, más aún, el producto de dos funciones Fréchet diferenciables es una función Fréchet diferenciable.

Lema 1.2.3 *Supongamos que X, Y y Z son espacios de Banach y $g : X \to Y$ y $f : Y \to Z$ son Fréchet diferenciables. Entonces la derivada de la composición está dada por la regla de la cadena*

$$(f \circ g)'(x) = f'(g(x))g'(x).$$

Si la derivada de Gâteaux existe y es continua en el siguiente sentido, entonces las dos nociones coinciden

Proposición 1.2.4 *Si f es Gâteaux diferenciable en una vecindad abierta U de x y $Df(x)$[1] es continua, entonces f es Fréchet diferenciable en x.*

La noción de Gâteaux diferenciable y Fréchet diferenciable también coinciden si f es Lipschitz y X es un espacio de dimensión finita:

[1]En el sentido de que $Df : U \to \mathcal{B}(X,Y)$ es continuo en x así que $\lim_{\tilde{x} \to x} \|Df(x) - Df(\tilde{x}\| = 0$. En otras palabras, la derivada depende continuamente en el punto x.

Proposición 1.2.5 *Supongamos que $f : X \to Y$ es una función Lipschitz de un espacio de Banach con dimensión finita a (posiblemente a un espacio de dimensión infinita) un espacio de Banach Y. Si f es Gâteaux diferenciable en algún punto x, entonces es Fréchet diferenciable en dicho punto.*

En el caso de dimensión infinita la historia es totalmente distinta. Hablando en términos generales, en tal situación existen resultados para Gâteaus diferenciabilidad de funciones Lipschitz, mientras que para funciones Fréchet diferenciables son inusuales y difíciles de demostrar. Por otro lado, en muchas aplicaciones es importante tener derivadas de Fréchet de f, ya que proporcionan una buena aproximación lineal y local de f a diferencia de las derivadas de Gâteaux.

Derivadas Parciales

En lo siguiente mencionamos algunas propiedades de las derivadas parciales.

Recordemos que $\mathcal{H}_1 \oplus \mathcal{H}_2$ es el espacio $\mathcal{H}_1 \times \mathcal{H}_2$ con la estructura lineal usual. Si $\mathcal{H}_1$ y $\mathcal{H}_2$ son espacios de Banach (Hilbert), entonces $\mathcal{H}_1 \oplus \mathcal{H}_2$ es también un espacio de Banach (Hilbert).

Notación (Derivadas Parciales) *Sea $f : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{F}$ diferenciable, denotamos las derivadas parciales de f como*

$$\begin{aligned} D_1 f(x_1, x_2) &= D(f(\cdot, x_2))(x_1) \in \mathcal{L}(\mathcal{H}_1, \mathcal{F}), \\ D_2 f(x_1, x_2) &= D(f(x_1, \cdot))(x_2) \in \mathcal{L}(\mathcal{H}_2, \mathcal{F}). \end{aligned}$$

Proposición 1.2.7 *Si $f : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{F}$ es diferenciable en (x_1, x_2), entonces $D_j f(x_1, x_2)$, $j = 1, 2$ existen y*

$$Df(x_1, x_2)(\xi_1, \xi_2) = D_1 f(x_1, x_2)\xi_1 + D_2 f(x_1, x_2)\xi_2. \tag{1.27}$$

Inversamente, si $U \subset \mathcal{H}_1 \oplus \mathcal{H}_2$ es abierto, y $D_1 f$ y $D_2 f$ existen y son continuas en U, entonces Df existe y se cumple (1.27).

Notación *Sean $\mathcal{H}$, $\mathcal{F}_1$ y $\mathcal{F}_2$ espacios de Banach, y consideremos $f : \mathcal{H} \to \mathcal{F}_1$, $g : \mathcal{H} \to \mathcal{F}_2$. Denotamos*

$$\begin{aligned} (f, g) : \mathcal{H} &\to \mathcal{F}_1 \oplus \mathcal{F}_2, \\ (f, g)(x) &= (f(x), g(x)). \end{aligned}$$

Lema 1.2.9 *Sean $f : \mathcal{H} \to \mathcal{F}_1$ y $g : \mathcal{H} \to \mathcal{F}_2$ diferenciables en x, entonces (f, g) es diferenciable en x y*

$$D(f, g)(x) = (Df(x), Dg(x)),$$

es decir,

$$D(f,g)(x)\xi = (Df(x)\xi, Dg(x)\xi).$$

Proposición 1.2.10 *Si $\beta : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{F}$ es bilineal y continuo, entonces β es diferenciable sobre $\mathcal{H}_1 \oplus \mathcal{H}_2$ y*

$$D\beta(x_1,x_2)(\xi_1,\xi_2) = \beta(\xi_1,x_2) + \beta(x_1,\xi_2).$$

Teorema 1.2.11 (Regla de Leibniz) *Sean $u : \mathcal{H} \to \mathcal{F}_1$, $v : \mathcal{H} \to \mathcal{F}_2$ diferenciables en $x \in \mathcal{H}$. Si $\beta : \mathcal{F}_1 \oplus \mathcal{F}_2 \to \mathcal{G}$ es bilineal y continuo, y si $f : \mathcal{H} \to \mathcal{G}$ es definido como*

$$f(y) = \beta(u(y), v(y)),$$

para todo $y \in \mathcal{H}$, entonces f es diferenciable en x y

$$Df(x)\xi = \beta(Du(x)\xi, v(x)) + \beta(u(x), Dv(x)\xi).$$

Ejemplo 1.2.1 *Sea $\mathcal{H}$ un espacio de Hilbert con producto interno $\langle\cdot,\cdot\rangle$. Consideremos $f : \mathcal{H} \to \mathbb{R}$ definida como*

$$f(x) = \frac{1}{2}\|x\|^2.$$

Más aún, sean $\beta : \mathcal{H} \oplus \mathcal{H} \to \mathbb{R}$ y $U : \mathcal{H} \to \mathcal{H} \oplus \mathcal{H}$ dados por

$$\begin{aligned} \beta(x,y) &= \frac{1}{2}\langle x, y\rangle, \\ U &= \langle I, I\rangle, \end{aligned}$$

donde I la matriz identidad. Entonces

$$f = \beta \circ U.$$

De esta manera,

$$\begin{aligned} Df(x)\xi &= D\beta(U(x)) \circ DU(x)\xi \\ &= D\beta(U(x)) \circ (DI(x)\xi, DI(x)\xi) \\ &= D\beta(U(x)) \circ (\xi,\xi) \\ &= \beta(\xi,x) + \beta(x,\xi) \\ &= \frac{1}{2}\langle \xi, x\rangle + \frac{1}{2}\langle x, \xi\rangle \\ &= \langle \xi, x\rangle. \end{aligned}$$

Gradiente por el Método Adjunto

Dual de un Espacio Normado

Definición 1.2.12 *Sean V, W espacios normados sobre $\mathbb{R}$. Un mapeo $\mathcal{A} : V \to W$ es lineal si para todos $u, v \in V$ y para todo $\lambda, \mu \in \mathbb{R}$ se cumple*

$$\mathcal{A}(\lambda u + \mu v) = \lambda \mathcal{A}u + \mu \mathcal{A}v.$$

$\mathcal{A}$ es continuo si $v_n \to v$, entonces

$$\mathcal{A}v_n \to \mathcal{A}v \text{ en } W.$$

A esta convergencia, se le conoce como convergencia fuerte.

Observemos que en espacios de dimensión infinita, un funcional lineal no es necesariamente continuo.

El conjunto de mapeos continuos de $V \to W$ es un espacio vectorial que denotamos como $\mathcal{L}(V, W)$. Es fácil demostrar que un mapeo lineal $\mathcal{A}$ de V a W es continuo si y sólo si existe una constante $M > 0$ tal que para todo $v \in V$

$$\|\mathcal{A}v\|_W \leq M \|v\|_V .$$

En vista de esta propiedad, podemos asociar a cada mapeo lineal continuo $\mathcal{A} : V \to W$ (que es un elemento de $\mathcal{L}(V, W)$) el número real

$$\|\mathcal{A}\|_{\mathcal{L}(V,W)} = \sup_{v \neq 0} \frac{\|\mathcal{A}v\|_W}{\|v\|_V},$$

y se puede demostrar que $\|\cdot\|_{\mathcal{L}(V,W)}$ es una norma en el espacio vectorial $\mathcal{L}(V, W)$. Para un demostración de estas propiedades puede consultarse [7]. Más aún, se puede demostrar la siguiente proposición.

Proposición 1.2.13 *Sea V un espacio normado, W un espacio de Banach. Entonces $\mathcal{L}(V, W)$ es un espacio de Banach.*

Esta proposición motiva la siguiente definición.

Definición 1.2.14 *Sea V un espacio normado. Definimos el espacio dual de V como el espacio de mapeos lineales continuos de V a $\mathbb{R}$, esto es $\mathcal{L}(V, \mathbb{R})$. El espacio dual de V es un espacio normado y lo denotamos como V^*, más aún, este espacio, es un espacio de Banach (debido a la proposición 1.2.13). A los elementos de $\mathcal{L}(V, \mathbb{R})$ son llamados funcionales.*

Para $\mathcal{L} \in V^*$, el valor de $\mathcal{L}$ en $v \in V$ lo denotamos por $\mathcal{L}v$, y la norma de $\mathcal{L}$ en V^* es definida como

$$\|\mathcal{L}\|_{V^*} = \sup_{v \neq 0} \frac{|\mathcal{L}v|}{\|v\|_V}.$$

A continuación daremos algunos ejemplo de espacios duales

Ejemplo 1.2.2 *Si $V = \mathbb{R}^n$, entonces es conocido que el dual V^* es isomorfo a $\mathbb{R}^n$.*

Ejemplo 1.2.3 *Sea $1 < p < \infty$, el espacio dual de $V = l^p$ es $V^* = l^q$, donde p y q se relacionan mediante la relación*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{1.28}$$

Para una demostración de esta afirmación, puede consultarse [7].

Ejemplo 1.2.4 *Sea Ω un conjunto abierto de $\mathbb{R}^n$ y p tal que $1 < p < \infty$. Entonces el dual de $V = \mathbb{L}^p(\Omega)$ es $V^* = \mathbb{L}^q(\Omega)$ donde p y q se relacionan mediante* (1.28).

Ya que el espacio dual V^* de un espacio vectorial es un espacio de Banach, podemos definir el dual V^{**} de V^*. Es evidente que V^{**} es un espacio de Banach y es llamado bidual de V.

Definición 1.2.15 *Decimos que un espacio de Banach es reflexivo si $V^{**} \equiv V$.*

Ejemplo 1.2.5 *El bidual de $\mathbb{R}^n$ (respectivamente de l^p, $\mathbb{L}^p(\Omega)$) es isomorfo a $\mathbb{R}^n$ (respectivamente con l^p, $\mathbb{L}^p(\Omega)$).*

Observemos que si V es un espacio de Banach reflexivo, entonces V^* es un espacio de Banach reflexivo.

Espacios de Hilbert

Definición 1.2.16 (Espacio de Hilbert) *Un espacio de Hilbert $\mathcal{H}$ es un espacio vectorial equipado con un producto escalar $\langle \cdot, \cdot \rangle$ tal que $\mathcal{H}$ es completo con la norma*

$$\|u\| = \langle u, u \rangle^{1/2}.$$

Ejemplo 1.2.6 *$\mathbb{R}^n$ con el producto escalar*

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i,$$

donde $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$.

Ejemplo 1.2.7 *Sea Ω un subconjunto abierto de $\mathbb{R}^n$ y consideremos el espacio de funciones cuadrado integrables, $\mathbb{L}^2(\Omega)$, con producto interno*

$$\langle u, v\rangle = \int_{\Omega} uvdx.$$

Ejemplo 1.2.8 *Consideremos el espacio de funciones continuas en el intervalo $[a, b]$, $\mathcal{C}[a, b]$, como en el ejemplo anterior, consideremos el producto interno*

$$\langle u, v\rangle = \int_{a}^{b} uvdx.$$

Sin embargo, $\mathcal{C}[a, b]$ no es un espacio completo para la topología asociada con la norma inducida con este producto interno. Así, $\mathcal{C}[a, b]$ no es un espacio de Hilbert.

Ejemplo 1.2.9 *Sea Ω un subconjunto de $\mathbb{R}^n$ abierto, y consideremos el espacio*

$$\mathbb{H}^1(\Omega) = \left\{ v : \ v \in \mathbb{L}^2(\Omega), \quad \frac{\partial v}{\partial x_i} \in \mathbb{L}^2(\Omega), \quad i = 1, \ldots, n \right\}$$

equipado con el producto interno

$$\langle u, v\rangle = \int_{\Omega} \left(uv + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} \right) dx.$$

Ya que un espacio de Hilbert $\mathcal{H}$ es un espacio de Banach, podemos definir su espacio dual $\mathcal{H}^*$ y su bidual $\mathcal{H}^{**}$.

El siguiente teorema nos muestra como actúa la reflexividad para espacios de Hilbert.

Teorema 1.2.17 (Teorema de Representación de Riesz) *Sea $\mathcal{H}$ un espacio de Hilbert y $\mathcal{L} \in \mathcal{H}^*$ un funcional lineal y continuo sobre $\mathcal{H}$. Entonces existe un único elemento $u \in \mathcal{H}$ tal que para todo $v \in \mathcal{H}$*

$$\mathcal{L}(v) = \langle u, v\rangle$$

y

$$\|\mathcal{L}\|_{\mathcal{H}^*} = \|u\|_{\mathcal{H}}.$$

Inversamente, podemos asociar a cada $u \in \mathcal{H}$ un funcional lineal continuo $\mathcal{L}_u$ tal que para todo $v \in \mathcal{H}$,

$$\mathcal{L}_u(v) = \langle u, v\rangle.$$

Una demostración de este teorema puede ser consultada en [7].

Una consecuencia inmediata de este teorema es el hecho de que $\mathcal{H}^{**}$ puede se identificado con $\mathcal{H}$, en otras palabras, todo espacio de Hilbert $\mathcal{H}$ es reflexivo.

Definición 1.2.18 (Gradiente de un funcional) *Sea $\mathcal{H}$ un espacio de Hilbert equipado con el producto escalar $\langle \cdot, \cdot \rangle$. Si $\mathcal{J}$ es Gâteaux diferenciable en $v \in \mathcal{H}$, y $\delta\mathcal{J}(v, \varphi)$ es lineal y continuo con respecto a φ, entonces existe un elemento $\mathcal{J}'(v) \in \mathcal{H}$ tal que para todo $\varphi \in \mathcal{H}$*

$$\delta\mathcal{J}(v, \varphi) = \left\langle \mathcal{J}'(v), \varphi \right\rangle,$$

llamamos $\mathcal{J}'(v)$ el gradiente de $\mathcal{J}$ en v.

Ejemplo 1.2.10 *Si $\mathcal{J}(v_1, v_2, \ldots, v_n)$ es una función de $\mathbb{R}^n$ a $\mathbb{R}$, diferenciable en el sentido usual, entonces*

$$\mathcal{J}'(v) = \left(\frac{\partial \mathcal{J}}{\partial v_1}, \frac{\partial \mathcal{J}}{\partial v_2}, \ldots, \frac{\partial \mathcal{J}}{\partial v_n} \right)^T.$$

De hecho, para todo $\varphi \in \mathbb{R}^n$

$$\delta\mathcal{J}(v, \varphi) = [\mathcal{J}'(v)]^T \cdot \varphi.$$

Ejemplo 1.2.11 *Sea $\mathcal{J}(v)$ el funcional de $\mathbb{L}^2[a, b] \to \mathbb{R}$ definido como*

$$\mathcal{J}(v) = \int_a^b [v(x)]^2 dx.$$

tenemos que para todo $\varphi \in \mathbb{L}^2[a, b]$

$$\frac{\mathcal{J}(v + \theta\varphi) - \mathcal{J}(v)}{\theta} = \int_a^b \varphi \left[2v + \theta\varphi\right] dx \to 2 \int_a^b v\varphi dx$$

y por lo tanto $\delta\mathcal{J}(v, \varphi)$ es lineal y continuo con respecto a φ, más aún,

$$\delta\mathcal{J}(v, \varphi) = \int_a^b 2v\varphi dx = \langle 2v, \varphi \rangle.$$

Por lo tanto,

$$\mathcal{J}'(v) = 2v.$$

Ejemplo 1.2.12 *Consideremos el funcional* $\mathcal{J} : \mathbb{H}^1[a,b] \to \mathbb{R}$ *definido como*

$$\mathcal{J}(v) = \int_a^b v^2(x)dx.$$

De manera similar ala ejemplo anterior,

$$\delta\mathcal{J}(v,\varphi) = 2\int_a^b v\varphi dx,$$

que es en efecto, una forma lineal y continua con respecto a φ*.*

Ya que el producto escalar en $\mathbb{H}^1[a,b]$ *se define mediante*

$$\langle v, w\rangle = \int_a^b \left(vw + \frac{\partial v}{\partial x}\frac{\partial w}{\partial x} \right) dx$$

el gradiente de $\mathcal{J}$ *no es igual a* $2v$ *como en el ejemplo anterior.*

Esta situación nos demuestra que para determinar el gradiente, es esencial especificar en cual espacio el funcional $\mathcal{J}$ *está definido.*

Ejemplo 1.2.13 *Sean* $\mathcal{K}, \mathcal{F} : \mathbb{L}^2(a,b) \to \mathbb{L}^2(a,b)$ *operadores lineales y consideremos la ecuación diferencial parcial*

$$u_t - \mathcal{F}u = \mathcal{K}p$$

donde p *es un parámetro inherente al problema. Más aún, consideremos el operador de mínimos cuadrados* $\mathcal{J} : \mathbb{L}^2(a,b) \to \mathbb{R}$

$$\mathcal{J}(p) = \frac{1}{2}\left\|\mathcal{M}u - \hat{u}\right\|^2.$$

que surge de manera natural en problemas de optimización. Cabe mencionar que

$$\mathcal{M} : \mathcal{C}(a,b) \times (0,T) \to \mathbb{R}^{M\times N}$$

es un operador de observación.

Nuestro fin es hallar una expresión para $\nabla\mathcal{J}(p)$*, para tal propósito consideremos el operador*

$$\mathcal{L}(p,u,\lambda) = \frac{1}{2}\left\|\mathcal{M}u - \hat{u}\right\|^2 + \langle \lambda, u_t - \mathcal{L}u - \mathcal{K}p\rangle.$$

La derivada del operador $\mathcal{L}$ está dada por

$$D\mathcal{L}(p,u,\lambda)(\xi,\eta) = D_1\mathcal{L}(p,u,\lambda)\xi + D_2\mathcal{L}(p,u,\lambda)\eta.$$

A continuación, calcularemos estas derivadas

$$\begin{aligned} D_1\mathcal{L}(p,u,\lambda)\xi &= \langle \lambda, -\mathcal{K}\xi \rangle, \\ D_2\mathcal{L}(p,u,\lambda)\eta &= \langle \eta, \mathcal{M}^*(\mathcal{M}u - \hat{u}\rangle + \langle \lambda, \eta_t - \mathcal{F}\eta \rangle. \end{aligned}$$

Sea

$$W(p) = (p, u(p)),$$

donde $u(p)$ resuelve el problema de Cauchy. Así,

$$\mathcal{J}(p) = \mathcal{L}(W(p)).$$

Al aplicar la regla de cadena, obtenemos

$$\begin{aligned} D\mathcal{J}(p)\xi &= D\mathcal{L}(W(p))DW(p)\xi \\ &= D\mathcal{L}(W(p))(\xi, Du(p)\xi) \\ &\equiv D\mathcal{L}(W(p))(\xi, U). \end{aligned}$$

De esta manera,

$$D\mathcal{J}(p)\xi = \langle \lambda, -\mathcal{K}\xi \rangle + \langle \lambda, U_t - \mathcal{F}U \rangle + \langle U, \mathcal{M}^*(\mathcal{M}u - \hat{u}) \rangle.$$

Al usar integración por partes podemos reescribir la expresión anterior como

$$\begin{aligned} D\mathcal{J}(p)\xi &= \langle \xi, -\mathcal{K}^*\lambda \rangle \\ &+ \langle U, -\lambda_t - \mathcal{F}^*\lambda + \mathcal{M}^*(\mathcal{M}u - \hat{u}) \rangle. \end{aligned}$$

Por lo tanto, si λ es solución de

$$\lambda_t + \mathcal{F}^*\lambda = \mathcal{M}^*(\mathcal{M}u - \hat{u}),$$

entonces

$$D\mathcal{J}(p)\xi = \langle \xi, -\mathcal{K}^*\lambda \rangle.$$

Al usar integración por partes y el teorema de representación de Riesz,

$$\nabla\mathcal{J}(p) = -\int_0^T \mathcal{K}^*\lambda(x,t)dt.$$

Capítulo 2

Formulación del Problema de Estimación de Batimetría en las Ecuaciones de Saint-Venant

2.1. Flujo Unidimensional en Aguas Someras

En esta sección describiremos la dinámica de las ondas de agua superficiales bajo el supuesto de aguas someras (aguas poco profundas). Esta dinámica está regida por las ecuaciones de Saint-Venant, que es un modelo habitualmente usado para describir el flujo en ríos, canales o áreas costeras. La principal suposición en el modelo es que la escala de longitud vertical z es mucho menor que las escalas horizontales x, y. Más aún, suponemos que el movimiento no depende de la variable y. Observemos la figura 2.1.

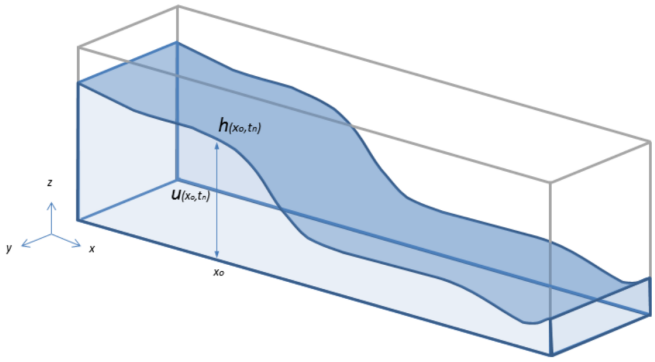


Figura 2.1: Flujo unidimensional en aguas someras.

El nivel de agua en reposo corresponde a $z = 0$. Supongamos que el fondo es plano e impermeable en $z = -H$. La superficie libre es $z = \eta(x, t)$. De esta manera, la altura de la superficie del agua con respecto al fondo es $h(x, t) = \eta(x, t) + H$. En el interior del fluido, los componentes de la velocidad $\mathbf{v}$ son (u, w). En la superficie, nos referimos a estas componentes como (U, W).

Balance de masa

Consideremos la masa entre dos líneas verticales en $x = a$ y $x = b$ como mostramos en la figura 2.2.

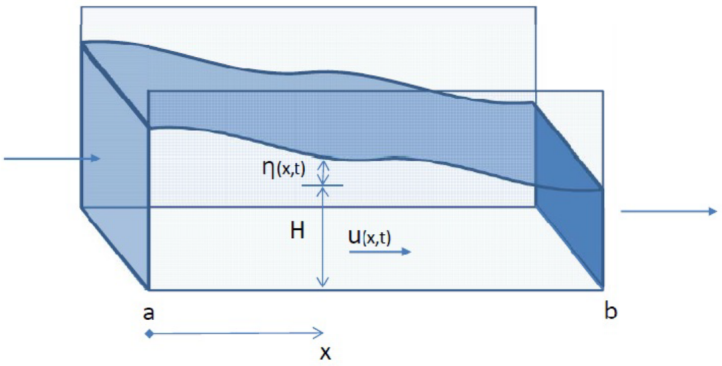


Figura 2.2: Dominio para balance.

De esta manera, la masa en la región es

$$m_{a,b} = \rho \int_a^b H(x,t)dx = \rho H(b-a) + \rho \int_a^b \eta(x,t)dx.$$

Los cambios de masa por unidad de tiempo se deben al flujo de agua por las líneas verticales. El flujo de agua $Q(x,t)$ a través de una línea en x está dado por

$$Q(x,t) = \rho \int_{-H}^{\eta(x,t)} u(x,z,t)dz.$$

Usando la conservación de masa

$$\frac{dm_{a,b}(t)}{dt} = -[Q(b,t) - Q(a,t)],$$

obtenemos la ecuación de balance

$$\partial_t \eta(x,t) = -\partial_x \int_{-H}^{\eta(x,t)} u(x,z,t)dx.$$

Si calculamos la derivada en el lado derecho, obtenemos

$$\partial_x \int_{-H}^{\eta(x,t)} u(x,z,t)dz = u(x,\eta(x,t),t)\partial_x \eta(x,t) + \int_{-H}^{\eta(x,t)} \partial_x u(x,z,t)dz.$$

Bajo el supuesto de que el fluido es incompresible se satisface

$$\nabla \cdot \mathbf{v} = \partial_x u + \partial_z w = 0.$$

De esta manera, obtenemos la simplificación

$$\begin{aligned} \int_{-H}^{\eta(x,t)} \partial_x u(x,z,t)dz &= -\int_{-H}^{\eta(x,t)} \partial_z w(x,z,t)dz \\ &= -w(x,\eta(x,t),t) + w(x,-H,t) \\ &= -W(x,t). \end{aligned}$$

Hemos usado el hecho que $w(x,-H,t) = 0$ pues el fondo es impermeable. Por lo tanto, en la superficie libre se satisface la ecuación

$$\partial_t \eta = -U\partial_x \eta + W.$$

Esta ecuación la podemos escribir de manera equivalente

$$\partial_t \eta = (U,W) \cdot (-\partial_x \eta, 1).$$

Observemos que esta ecuación es válida sin el supuesto de aguas someras. Al usar esta relación podemos considerar que u no depende de z y de la ecuación de balance se desprende

$$\partial_t \eta + \partial_x[(\eta + H)u] = 0,$$

o de manera equivalente

$$\partial_t h + \partial_x(hu) = 0.$$

Balance de Momento

De la segunda ley de Newton sabemos que la derivada del momento $\mathbf{p}$ es igual a la suma de fuerzas,

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}.$$

En aguas someras el momento en x es

$$M(x,t) = \int_{-H}^{\eta} \rho u(x,t)dz = \rho h u(x,t).$$

La fuerza incluye el flujo de momento

$$Q(x,t) = \int_{-H}^{\eta} \rho u^2(x,t)dz = \rho h u^2(x,t),$$

y la presión a lo largo de la vertical es

$$P(x,t) = \int_{-H}^{\eta} p(x,z,t)dz,$$

donde $p(x,z,t)$ es a presión local. Como antes podemos obtener la ecuación de balance

$$\partial_t M + \partial_x Q + \partial_x P = 0.$$

De esta manera,

$$\partial_t(\rho h u) = -\partial_x \left\{ \rho h u^2 + \int_{-H}^{\eta} p(x,z,t)dz \right\}.$$

En aguas someras, la presión local se aproxima por la presión hidrostática

$$p = \rho g(\eta - z) + p_{\text{atm}},$$

donde g es la constante gravitacional, y p_{atm} se supone igual a cero. Entonces,

$$\partial_t(hu) = -\partial_x \left\{ hu^2 + \frac{1}{2}gh^2 \right\}.$$

Al usar la ecuación de continuidad

$$\partial_t u + u\partial_x u + g\partial_x h = 0,$$

podemos establecer

$$\partial_t u + \partial_x \left\{ \frac{1}{2}u^2 + gh \right\}.$$

Finalmente, podemos escribir ambas ecuaciones de balance en su forma vectorial

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ u \end{bmatrix} + \frac{\partial}{\partial x}\begin{bmatrix} hu \\ \frac{1}{2}u^2 + gh \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

2.2. Batimetría no Nula

En nuestro planteamiento, el sistema de ecuaciones de aguas someras, pueden ser escritas como un sistema de ecuaciones diferenciales parciales hiperbólicas con un término

fuente más general como

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ u \end{bmatrix} + \frac{\partial}{\partial x}\begin{bmatrix} hu \\ \frac{1}{2}u^2 + gh \end{bmatrix} = \begin{bmatrix} 0 \\ F(\Theta, h, u) \end{bmatrix}, \quad x \in [a, b], \quad t \geq 0.$$

Aquí h es la profundidad, u la velocidad, y g es la constante gravitacional. El parámetro Θ corresponde a la situación que se este modelando.

Para obtener un problema bien planteado, requerimos condiciones iniciales y de frontera adecuadas:

$$h(x, 0) = h_0(x), \quad u(x, 0) = u_0(x), \quad x \in (a, b),$$

y

$$h(a, t) = h_a(t), \quad u(a, t) = u_a(t), \quad t \in (0, T),$$

donde las funciones $h_0(x)$, $u_0(x)$, $h_a(t)$ y $u_a(t)$ serán especificadas cuando efectuemos simulaciones numéricas.

En el caso en que estemos tomando en cuenta la batimetría y el coeficiente de Manning, el término fuente es

$$F(\Theta, h, u) = -gB_x - \kappa h^{1-\eta} u|u|.$$

Así $\Theta = (B, k)$. En esta situación, $B(x)$ es la topografía, κ es el coeficiente de fricción de Manning, y η es un parámetro que usualmente se toma como $7/3$.

Al considerar solo el término de batimetría, es decir, el fondo está descrito por una función $B(x)$, entonces el sistema de ecuaciones diferenciales parciales es

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ u \end{bmatrix} + \frac{\partial}{\partial x}\begin{bmatrix} hu \\ \frac{1}{2}u^2 + gh \end{bmatrix} = \begin{bmatrix} 0 \\ -gB_x \end{bmatrix}, \quad x \in [a, b], \quad t \geq 0. \tag{2.1}$$

El lado izquierdo de este sistema es un operador de transporte, correspondiente al flujo de un fluido ideal en un canal plano, sin fricción, lluvia o infiltración. Este es el modelo propuesto por Saint-Venant en [35], y contiene varias propiedades importantes del flujo. A fin de enfatizar estas propiedades, primero reescribimos las ecuaciones unidimensionales en su forma vectorial:

$$\partial_t W + \partial_x F(W) = G,$$

donde

$$W = \begin{bmatrix} h \\ u \end{bmatrix}, \quad F'(W) = \begin{bmatrix} u & h \\ g & u \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ -gB_x(x) \end{bmatrix},$$

con $F(W)$ el flujo de la ecuación. El transporte es más claro en la siguiente forma no conservativa, donde $A(W) = F'(W)$

$$\partial_t W + A(W)\partial_x W = G.$$

Observemos que cuando $h > 0$, la matriz $A(W)$ es diagonalizable con valores propios

$$\lambda_1(W) = u - \sqrt{gh} < u + \sqrt{gh} = \lambda_2(W).$$

Esta importante propiedad es llamada hiperbolicidad estricta. Los valores propios son en efecto velocidades de las ondas superficiales del fluido. Observemos que los valores propios coinciden cuando $h = 0$, esto es para zonas secas. En este caso, el sistema deja de ser hiperbólico, y esto induce dificultades a nivel teórico y numérico. Diseñar esquemas numéricos que preserven la no negatividad de h son importantes en este contexto.

De estas fórmulas recuperamos una clasificación útil de los flujos, basados en los valores relativos de la velocidad del fluido, u y de las ondas, $\sqrt{gh}$. En efecto, si $|u| < \sqrt{gh}$ las velocidades características tienen signos opuestos, y la información se propaga hacia arriba y abajo del flujo, a la cual llamamos subcritical o fluvial. Por otro lado, cuando $|u| > \sqrt{gh}$, el flujo es super critico, o torrencial, toda la información va hacia abajo. Un régimen transcrito existe cuando algunas partes del flujo son subcríticos o supercríticos.

Al hacer el cambio de variable $q = hu$, podemos escribir el sistema (2.1) como

$$\begin{cases} \partial_t h + \partial_x q = 0 \\ \partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{gh^2}{2} \right) = -ghB_x(x) \end{cases} \qquad x \in [a, b], \quad t \geq 0. \tag{2.2}$$

Por un momento, nos centraremos en soluciones estacionarias del sistema (2.3), es decir, soluciones que satisfacen

$$\partial_t h = 0 \quad \text{y} \quad \partial_t q = 0.$$

Reemplazando esta relación en las ecuaciones de aguas someras, obtenemos

$$\begin{aligned} \partial_x q &= 0, \\ \partial_x \left(\frac{q^2}{h} + \frac{gh^2}{2} \right) &= -ghB_x(x). \end{aligned}$$

Supongamos que $h \neq 0$, entonces

$$\begin{cases} q = q_0, \\ B_x(x) = \frac{1}{gh}\left(\frac{q^2}{h^2} - gh\right)\partial_x h. \end{cases} \tag{2.3}$$

En el caso de soluciones regulares, obtenemos la relación

$$\frac{q_0^2}{2gh^2(x)} + h(x) + B(x) = C, \tag{2.4}$$

que nos proporciona un vínculo entre la batimetría y la altura.

2.3. Formulación del Problema

Un problema no abordado en la literatura, es la determinación de la batimetría usando mediciones de velocidad. Recientemente se han diseñado artefactos para ello como en [5].

Supongamos que medimos la velocidad u en los puntos (x_j, t_n), $j = 1, 2 \ldots, M$, $n = 1, 2, \ldots, N$, es decir, $u(x_j, t_n) \approx \hat{u}_{j,n}$.

El problema de identificación es estimar la batimetría $B(x)$ a partir de un conjunto de observaciones de la velocidad. Para esto consideremos el problema no lineal en sentido de mínimos cuadrados, es decir, si definimos el funcional $\mathcal{J} : \mathbb{L}^2(a, b) \to \mathbb{R}$ dado por

$$\mathcal{J}(B) = \frac{1}{2}\sum_{j,n}\left[u(x_j, t_n; B) - \hat{u}_{j,n}\right]^2.$$

El problema de nuestro interés es minimizar $\mathcal{J}$ restringido a que h y u son soluciones de las ecuaciones de aguas someras con condiciones iniciales y de frontera adecuadas.

Capítulo 3

El Método de Descenso Continuo

3.1. Algoritmo de Optimización

En esta sección describiremos el algoritmo de descenso continuo para resolver el problema de optimización

$$\text{mín}\, \mathcal{J}(B)$$

sujeto a que $h(B)$ y $u(B)$ son soluciones de las ecuaciones de aguas someras con batimetría B.

Para encontrar el mínimo del funcional $\mathcal{J}$, usaremos el método de descenso continuo en el cual construiremos una sucesión B_k, $k = 1, 2, \ldots$ con la forma

$$B_{k+1} = B_k + \alpha_k p_k$$

donde α_k es el tamaño del paso (que tan lejos descendemos) y p_k es una dirección de descenso. La elección de ambos factores determinará la eficiencia del algoritmo. El cálculo del tamaño exacto del paso puede ser costoso computacionalmente. En realidad, lo más importante es buscar la dirección de descenso correcta, de esta manera, un tamaño razonable de paso, α_k, no necesariamente exacto, durante cada iteración será suficiente. Para este propósito, usaremos un método llamado *búsqueda de linea.*

Para encontrar el mínimo de $\mathcal{J}$, intentaremos buscar a lo largo de la dirección de descenso p_k con un paso ajustable α_k tal que

$$\psi(\alpha_k) = \mathcal{J}(B_k + \alpha_k p_k)$$

disminuya lo más posible. Esto da a lugar a un problema de optimización escalar que puede ser resuelto, por ejemplo, mediante el método de razón dorada (Golden Section Search), para más detalles sobre este procesos se puede consultar [2] y [17].

En cualquier punto B_k donde el gradiente del operador $\mathcal{J}$ es distinto de cero, el negativo del gradiente, $-\nabla\mathcal{J}(B_k)$ nos proporciona la dirección optima de mayor decrecimiento ($\mathcal{J}$ disminuya más rápidamente a lo largo de la dirección del gradiente negativo que de cualquier otra), de esta manera es natural escoger $p_k = -\nabla\mathcal{J}(B_k)$.

Observemos que

$$\psi'(\alpha) = -\nabla\mathcal{J}(B_k - \alpha\nabla\mathcal{J}(B_k)) \cdot \nabla\mathcal{J}(B_k),$$

de esta manera

$$\psi'(0) = -\left\|\nabla\mathcal{J}(B_k)\right\|^2 < 0.$$

Por lo tanto, existe $\alpha^* > 0$ tal que $\psi(\alpha^*) < \psi(0)$. En otras palabras, el método de descenso continuo genera una sucesión monótona

$$\mathcal{J}(B_{k+1}) < \mathcal{J}(B_k),$$

esto es, el algoritmo garantiza la progresión hacia un mínimo en cada iteración. Como B_{k+1} se obtiene mediante la optimización de ψ, entonces

$$\psi'(\alpha_k) = -\nabla\mathcal{J}(B_{k+1}) \cdot \nabla\mathcal{J}(B_k) = 0,$$

es decir, los gradientes de iteraciones consecutivas son ortogonales. Más aún, el gradiente siempre es ortogonal a los conjuntos de nivel del funcional $\mathcal{J}$. De esta manera, las iteraciones se desplazan en zigzag hasta alcanzar el mínimo (ver figura 3.1) o cumplirse una condición de paro.

En el caso de que en este algoritmo tengamos $\mathcal{J}(B_k) = 0$ hemos llegado a un punto óptimo, esta situación raramente se da en procesos computacionales, por esto es plausible imponer la condición de paro

$$\|\mathcal{J}(B_k)\| < \varepsilon,$$

donde $\varepsilon > 0$ es una tolerancia dada. Una segunda condición de paro admisible es controlar el número de iteraciones.

Para un planteamiento formal de este proceso de optimización, así como un análisis de la convergencia se puede consultar [6], [17] y [44].

En el algoritmo 1 resumimos este proceso.

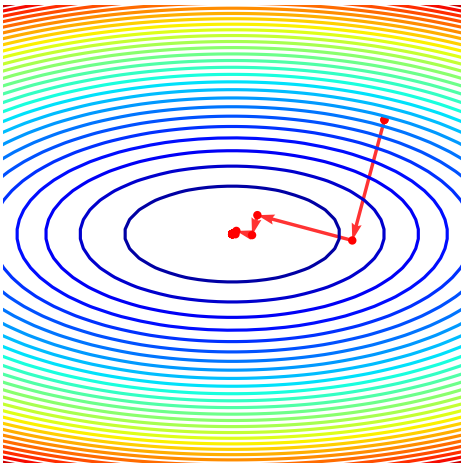


Figura 3.1: Método de descenso continuo.

Algorithm 1: Método de Descenso Continuo

B_0: Aproximación inicial.
ε: Tolerancia de convergencia.
$N_{\max}$: Máximo número de iteraciones.
$k \leftarrow 0$
while $\|\nabla \mathcal{J}(B_k)\| > \varepsilon$ **and** $k \leq N_{\max}$ **do**
  $p_k = -\nabla \mathcal{J}(B_k)$.
  Calcular $\alpha_k = \min \mathcal{J}(B_k - \alpha \nabla \mathcal{J}(B_k))$.
  $B_{k+1} = B_k + \alpha_k p_k$.
  $k \leftarrow k + 1$.
end while

En el teorema 3.1.2 establecemos condiciones necesarias para la convergencia del método de descenso continuo.

Definición 3.1.1 *Una función continua $f(\mathbf{x})$ definida en todo $\mathbb{R}^n$ es coercitiva si*

$$\lim_{\|\mathbf{x}\| \to \infty} f(\mathbf{x}) = \infty.$$

Esto significa que para cualquier constante M existe un número positivo R_M tal que $f(\mathbf{x}) \geq M$ siempre que $\|\mathbf{x}\| \geq R_M$. En particular, los valores de $f(\mathbf{x})$ no pueden permanecer acotados en un conjunto $A \subset \mathbb{R}^n$ que no es acotado.

Teorema 3.1.2 *Supongamos que f es una función coercitiva de clase $\mathcal{C}^1$. Si $\{\mathbf{x}_k\}$ es la sucesión generada por el método de descenso continuo a partir de una aproximación inicial $\mathbf{x}_0$, entonces alguna subsucesión de $\{\mathbf{x}_k\}$ es convergente. El límite de cualquier subsucesión convergente es un punto crítico de f.*

Para una demostración de este enunciado, puede revisarse [32].

Nuestro problema es encontrar una forma eficiente de evaluar el gradiente del operador $\mathcal{J}$, lo cual discutiremos en la siguiente sección.

3.2. Gradiente por el Método del Estado Adjunto

A continuación presentaremos un análisis que establece una expresión para $\nabla\mathcal{J}(B)$ suponiendo la diferenciabilidad Fréchet del funcional.

Sea el $\mathcal{M}$ el operador lineal de observación

$$\mathcal{M} : \mathcal{C}(\Omega \times (0,T)) \subset \mathbb{L}^2(\Omega \times (0,T)) \to \mathbb{R}^{M\times N}$$

dado por

$$\mathcal{M}u = \{u(x_j, t_n)\} \in \mathbb{R}^{M\times N},$$

y consideremos el Lagrangiano

$$\mathcal{L}(B,h,u,\lambda,\mu) = \frac{1}{2}\left\|\mathcal{M}u - \hat{u}\right\|^2 + \left\langle \begin{pmatrix}\lambda\\ \mu\end{pmatrix}, \begin{pmatrix} h_t + (uh)_x \\ u_t + \left(gh + \frac{1}{2}u^2\right)_x + gB_x \end{pmatrix} \right\rangle_{\mathbb{L}^2(\Omega\times[0,T])},$$

donde $\Omega = (a,b)$ y λ, μ son los multiplicadores de Lagrange.

En el siguiente teorema estableceremos una expresión para el gradiente del funcional $\mathcal{J}$.

Teorema 3.2.1 *Sean h y u soluciones de las ecuaciones de aguas someras para una batimetría $B(x)$. Supongamos que los multiplicadores de Lagrange son soluciones del sistema adjunto*

$$\begin{pmatrix}\lambda\\ \mu\end{pmatrix}_t + \begin{pmatrix} u & g \\ h & u\end{pmatrix}\begin{pmatrix}\lambda\\ \mu\end{pmatrix}_x = \begin{pmatrix} 0 \\ \mathcal{M}^*(\mathcal{M}u - \hat{u})\end{pmatrix}, \tag{3.1}$$

con condiciones terminales

$$\lambda(x,T) = 0, \quad \mu(x,T) = 0, \quad x \in \Omega$$

y de frontera

$$\lambda(b,t) = 0, \quad \mu(b,t) = 0, \quad t \in (0,T).$$

Entonces la derivada de Fréchet del funcional $\mathcal{J}$ es

$$D\mathcal{J}(B)\xi = \langle \xi, -g\mu_x \rangle \,, \tag{3.2}$$

de esta manera,

$$\nabla \mathcal{J}(B) = -g \int_0^T \mu_x(x,t)dt.$$

Demostración: La derivada del operador $\mathcal{L}$ está dada por

$$D\mathcal{L}(B,h,u,\lambda,\mu)(\xi,\eta,\zeta) = D_1\mathcal{L}(B,h,u,\lambda,\mu)\xi + D_2\mathcal{L}(B,h,u,\lambda,\mu)\eta + D_3\mathcal{L}(B,h,u,\lambda,\mu)\zeta.$$

A continuación, calcularemos estas derivadas

$$\begin{aligned}
D_1\mathcal{L}(B,h,u,\lambda,\mu)\xi &= \langle \mu, g\xi_x \rangle \,, \\
D_2\mathcal{L}(B,h,u,\lambda,\mu)\eta &= \langle \lambda, \eta_t + (u\eta)_x \rangle + \langle \mu, (g\eta)_x \rangle \,, \\
D_3\mathcal{L}(B,h,u,\lambda,\mu)\zeta &= \langle \zeta, \mathcal{M}^*(\mathcal{M}u - \hat{u}) \rangle + \langle \lambda, (h\zeta)_x \rangle + \langle \mu, \zeta_t + (u\zeta)_x \rangle \,.
\end{aligned}$$

Sea

$$W(B) = (B, h(B), u(B)),$$

donde $h(B)$, $u(B)$ resuelven las ecuaciones de aguas someras para B dada. Entonces

$$\mathcal{J}(B) = \mathcal{L}(W(B)).$$

Al aplicar la regla de la cadena, obtenemos

$$\begin{aligned}
D\mathcal{J}(B)\xi &= D\mathcal{L}(W(B))DW(B)\xi \\
&= D\mathcal{L}(W(B))(\xi, Dh(B)\xi, Du(B)\xi) \\
&\equiv D\mathcal{L}(W(B))(\xi, H, U).
\end{aligned}$$

De esta manera,

$$\begin{aligned}
D\mathcal{J}(B)\xi &= \langle \mu, g\xi_x \rangle \\
&+ \langle \lambda, H_t + (uH)_x \rangle + \langle \mu, (gH)_x \rangle \\
&+ \langle U, \mathcal{M}^*(\mathcal{M}u - \hat{u}) \rangle + \langle \lambda, (hU)_x \rangle + \langle \mu, U_t + (uU)_x \rangle \,.
\end{aligned}$$

Ya que la diferenciabilidad Fréchet implica Gâteaux diferenciabilidad (y ambas derivadas coinciden), entonces

$$Dh(B)\xi = \lim_{\varepsilon \to 0} \frac{h(B+\varepsilon\xi) - h(B)}{\varepsilon},$$

de manera similar,

$$Du(B)\xi = \lim_{\varepsilon \to 0} \frac{u(B+\varepsilon\xi) - u(B)}{\varepsilon}.$$

Estas expresiones implican que las funciones $H \equiv Dh(B)\xi$ y $U \equiv Du(B)\xi$ se anulan en los puntos donde tenemos condiciones iniciales y de frontera, en otras palabras,

$$H(x,0) = 0 = U(x,0) \quad \text{y} \quad H(a,t) = 0 = U(a,t).$$

Demostraremos la ecuación (3.2) para $\xi \in \mathcal{C}_c^\infty(a,b)$. Esto es suficiente ya que tal conjunto es denso en $\mathbb{L}^2(a,b)$.

Al usar integración por partes, las condiciones terminales y de frontera de λ, μ, H y U obtenemos:

$D_1\mathcal{L}(B,h,u,\lambda,\mu)\xi$

$$\begin{aligned}
\langle \mu, g\xi_x \rangle &= \int_0^T \int_a^b \mu g \xi_x dx dt \\
&= \int_0^T \left[\mu g \xi \Big|_a^b - \int_a^b \xi(\mu g)_x dx \right] dt \\
&= \langle \xi, -g\mu_x \rangle + \int_0^T \mu g \xi \Big|_a^b dt \\
&= \langle \xi, -g\mu_x \rangle .
\end{aligned}$$

$D_2\mathcal{L}(B,h,u,\lambda,\mu)H = \langle \lambda, H_t + (uH)_x \rangle + \langle \mu, (gH)_x \rangle$

$$\begin{aligned}
\langle \lambda, H_t \rangle &= \int_a^b \int_0^t \lambda H_t dt dx \\
&= \int_a^b \left[\lambda H \Big|_0^T - \int_0^T H \lambda_t dt \right] dx \\
&= \langle H, -\lambda_t \rangle .
\end{aligned}$$

$$\begin{aligned}
\langle \lambda, (uH)_x \rangle &= \int_0^T \int_a^b \lambda (uH)_x dx dt \\
&= \int_0^T \left[\lambda u H \Big|_a^b - \int_a^b \lambda_x u H dx \right] dt \\
&= \langle H, -u\lambda_x \rangle .
\end{aligned}$$

$$\begin{aligned}\langle \mu, (gH)_x \rangle &= \int_0^T \int_a^b \mu (gH)_x dx dt \\ &= \int_0^T \left[\mu g H \Big|_a^b - \int_a^b \mu_x g H dx \right] dt \\ &= \langle H, -g\mu_x \rangle .\end{aligned}$$

$$D_3\mathcal{L}(B,h,u,\lambda,\mu)U = \langle U, \mathcal{M}^*(\mathcal{M}u - \hat{u}) \rangle + \langle \lambda, (hU)_x \rangle + \langle \mu, U_t + (uU)_x \rangle$$

$$\begin{aligned}\langle \lambda, (hU)_x \rangle &= \int_0^t \int_a^b \lambda (hU)_x dx dt \\ &= \int_0^T \left[\lambda h U \Big|_a^b - \int_a^b \lambda_x h U dx \right] dt \\ &= \langle U, -h\lambda_x \rangle .\end{aligned}$$

$$\begin{aligned}\langle \mu, U_t \rangle &= \int_a^b \int_0^T \mu U_t dt dx \\ &= \int_a^b \left[\mu U \Big|_0^T - \int_0^T \mu_t U dt \right] dx \\ &= \langle U, -\mu_t \rangle .\end{aligned}$$

$$\begin{aligned}\langle \mu, (uU)_x \rangle &= \int_0^T \int_a^b \mu (uU)_x dx dt \\ &= \int_0^T \left[uU\mu \Big|_a^b - \int_a^b (uU\mu_x dx \right] dt \\ &= \langle U, -u\mu_x \rangle .\end{aligned}$$

Así,

$$\begin{aligned}D\mathcal{J}(B)\xi &= \langle \xi, -g\mu_x \rangle \\ &+ \langle H, -\lambda_t - u\lambda_x - g\mu_x \rangle \\ &+ \langle U, \mathcal{M}^*(\mathcal{M}u - \hat{u}) - h\lambda_x - \mu_t - u\mu_x \rangle .\end{aligned}$$

Entonces,

$$D\mathcal{J}(B)\xi = \langle \xi, -g\mu_x \rangle .$$

Finalmente, al usar integración por partes y el teorema de representación de Riesz, obtenemos que

$$\nabla \mathcal{J}(B) = -g \int_0^T \mu_x(x,t) dt,$$

como se quería. □

3.3. Aproximación del Operador Adjunto del Operador de Observación

Uno de los problemas a los que nos enfrentaremos al establecer un esquema numérico para resolver el sistema adjunto (3.1), es saber como actúa el operador $\mathcal{M}^*$ en el siguiente sentido: Sea $I \subset (a,b)$, φ una función suave y $\mathbf{V} \in \mathbb{R}^{M\times N}$ entonces nos interesa aproximar la integral

$$\int_I (\mathcal{M}^*\mathbf{V})(x,t)\varphi(x)dx.$$

Para tal propósito usaremos el teorema de diferenciación de Lebesgue:

Teorema 3.3.1 *Supongamos que $f \in \mathbb{L}^1_{\text{loc}}$. Entonces para casi toda x*

$$\lim_{r\to 0} \fint_{E_r} |f(y) - f(x)|dy = 0 \quad y \quad \lim_{r\to 0} \fint_{E_r} f(y)dy = f(x)$$

para toda familia $\{E_r\}_{r>0}$ que se comprime de una manera regular a x.

Una demostración de este teorema y de sus aplicaciones se puede consultar en [12].

Sea $\Delta t > 0$ y consideremos el conjunto $E_t = [t - \Delta t/2, t + \Delta t/2]$, en virtud del teorema 3.3.1

$$\begin{aligned}
\int_I (\mathcal{M}^*\mathbf{V})(x,t)\varphi(x)dx &= \int_a^b (\mathcal{M}^*\mathbf{V})(x,t)\varphi(x)\chi_I(x)dx \\
&\approx \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2}\int_a^b (\mathcal{M}^*\mathbf{V})(x,t)\varphi(x)\chi_I(x)dxdt \\
&= \frac{1}{\Delta t}\int_0^T\int_a^b (\mathcal{M}^*\mathbf{V})(x,t)\varphi(x)\chi_I(x)\chi_{E_t}(t)dxdt \\
&= \frac{1}{\Delta t}\left\langle (\mathcal{M}^*\mathbf{V})(x,t), \varphi(x)\chi_I(x)\chi_{E_t}(t)\right\rangle_{\mathbb{L}^2(\Omega\times[0,T])} \\
&= \frac{1}{\Delta t}\left\langle \mathbf{V}, \mathcal{M}(\varphi(x)\chi_I(x)\chi_{E_t}(t))\right\rangle_{\mathbb{R}^{m\times n}} \\
&= \frac{1}{\Delta t}\sum_{j,n} \mathbf{V}_{j,n}\left[\varphi(x_j)\chi_I(x_j)\chi_{E_t}(t_n)\right].
\end{aligned}$$

Capítulo 4

El Método de Galerkin Discontinuo

4.1. Leyes de Conservación Escalares

En aras de presentar las ideas fundamentales del método de Galerkin discontinuo para la solución numérica de ecuaciones diferenciales parciales, nos enfocaremos en desarrollar este método para resolver la ley de conservación escalar (4.1).

$$\begin{cases} u_t + f(u)_x = 0, & [a,b] \times (0,T), \\ u(x,0) = u_0(x) & x \in [a,b]. \end{cases} \tag{4.1}$$

Para discretizar en el espacio, procedemos de la siguiente forma: Para cada partición del intervalo $[a,b]$, $\left\{x_{i+1/2}\right\}_{i=0}^{M}$, definimos I_i como

$$I_i = (x_{i-1/2}, x_{i+1/2}), \quad i = 1, \ldots, M.$$

Sean

$$\Delta_i = x_{i+1/2} - x_{i-1/2}, \quad 1 \leq i \leq M; \quad h = \max_{1 \leq i \leq M} \Delta_i.$$

Asumiremos que la malla es regular, es decir, existe una constante $c > 0$ e independiente de h, tal que

$$\Delta_i \geq ch, \quad 1 \leq i \leq M.$$

Buscamos una aproximación $u^i(x,t)$ a $u(x,t)$ en el intervalo I_i tal que para cada tiempo $t \in [0,T]$, $u^i(\cdot,t)$ pertenece al espacio de dimensión finita

$$\mathbb{V}(I_i) := \left\{ v \in \mathbb{L}^2([a,b]) : \ v|_{I_i} \in \mathbb{P}^p(I_i) \right\},$$

donde $\mathbb{P}^p(I)$ denota el espacio de polinomios en I de grado a lo más p. A fin de determinar dicha solución aproximada, usamos una formulación débil obtenida de la siguiente manera: Primero, multiplicamos la ley de conservación escalar en (4.1) por una función suave y

arbitraria $\varphi(x)$ e integramos sobre I_i, así, después de integrar por partes

$$\int_{I_i} \partial_t u(x,t)\varphi(x)dx = \int_{I_i} f(u(x,t))\partial_x\varphi(x)dx - f(u(x,t))\varphi(x)\Big|_{\partial I_i}. \tag{4.2}$$

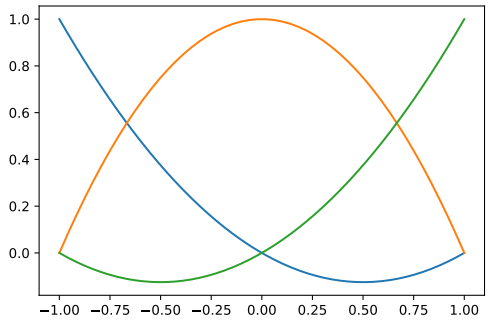
Ahora reemplacemos la función $\varphi(x)$, por una función prueba perteneciente al espacio $\mathbb{V}(I_i)$. Sea $x_{i-1/2} = x_0^i < x_1^i < \ldots < x_p^i = x_{i+1/2}$ una partición del elemento I_i. Entonces $u^i(x,t)$ es de la forma

$$u^i(x,t) = \sum_{k=0}^{p} u_k^i(t)N_k^i(x), \tag{4.3}$$

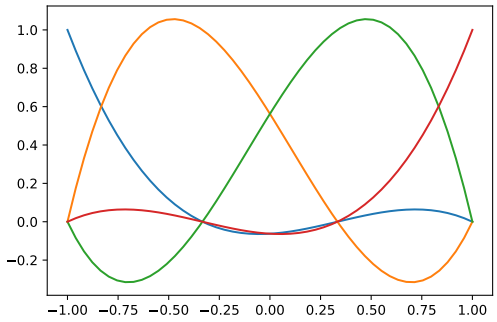
donde $N_k^i(x)$ son polinomios de Lagrange que satisfacen

$$N_k^i(x_l^i) = \delta_{kl}.$$

En la gráfica 4.1 mostramos los polinomios de Lagrange de grado $p = 2, 3$ en el intervalo $[-1, 1]$. En la gráfica 4.2 ilustramos la aproximación numérica del método Galerkin Discontinuo.



$p = 2$



$p = 3$

Figura 4.1: Polinomios de Lagrange en $[-1, 1]$.

De esta manera, para $l \in \{0, 1, \ldots, p\}$

$$\sum_{k=0}^{p} \frac{du_k^i(t)}{dt} \int_{I_i} N_k^i(x)N_l^i(x)dx = \int_{I_i} f\left(\sum_{k=0}^{p} u_k^i(t)N_k^i(x)\right)\partial_x N_l^i(x)dx - f\left(u^i(x,t)\right)N_l^i(x)\Big|_{\partial I_i}.$$

Ya que las aproximaciones $u^i(x,t)$ son discontinuas en los puntos $\{x_{i+1/2}, x_{i-1/2}\}$, debemos reemplazar a $u^i(x,t)$ en la frontera por un flujo numérico f^* que depende de los puntos extremos $x_{i+1/2}$, $x_{i-1/2}$ y la normal exterior n^-, es decir,

$$\begin{aligned} f_{i+1/2}^* &= f^*(u^i(x_{i+1/2}^-,t), u^i(x_{i+1/2}^+,t), n^-), \\ f_{i-1/2}^* &= f^*(u^i(x_{i-1/2}^-,t), u^i(x_{i-1/2}^+,t), n^-). \end{aligned}$$

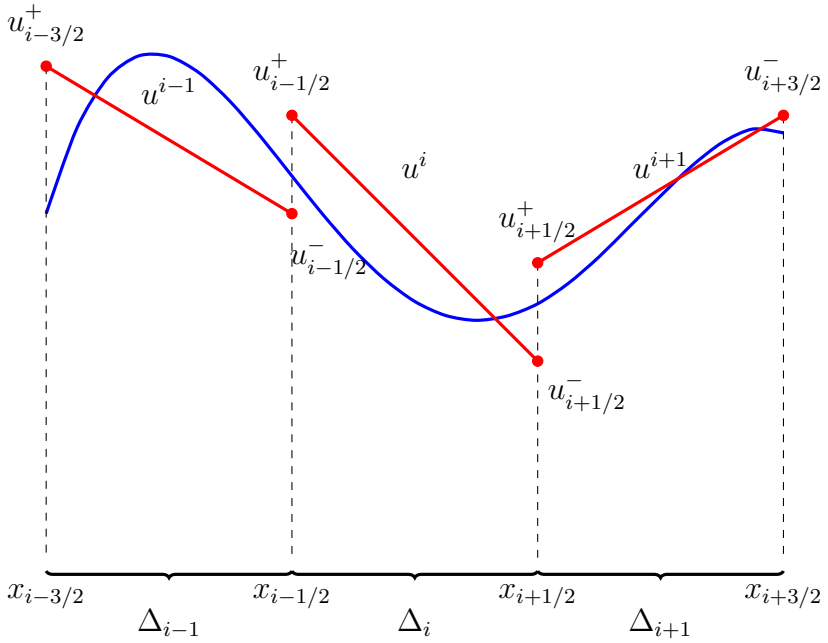


Figura 4.2: Reconstrucción de la solución u mediante el método de Galerkin Discontinuo.

Observemos que

$$\begin{aligned} f^*_{i+1/2} &= f^*(u^i_p(t), u^{i+1}_0(t), +1), \\ f^*_{i-1/2} &= f^*(u^i_0(t), u^{i-1}_p(t), -1). \end{aligned}$$

Usaremos los llamados *flujos monótonos* de los esquemas de diferencias finitas y volumen finito para resolver leyes de conservación, que satisfacen las condiciones:

- Consistencia: $f^*(u, u; n) = f(u \cdot n)$.
- Continuidad: $f^*(u^-, u^+; n)$ por lo menos es Lipschitz continuo con respecto a ambos argumentos u^-, u^+.
- Monotonicidad: $f^*(u^-, u^+; n)$ es una función no decreciente del primer argumento u^- y una función no creciente en el segundo argumento u^+.

Los ejemplos más conocidos de flujos numéricos que cumplen las condiciones anteriores son:

- El flujo de Lax-Friedrichs:

$$\begin{aligned} f^{LF}(a, b; n) &= \frac{1}{2}\left[n(f(a) + f(b)) - C(b - a)\right], \\ C &= \text{máx}\left\{|f'(s)| : \ \text{ínf}\, u_0(x) \leq s \leq \sup u_0(x)\right\}. \end{aligned}$$

- El flujo local de Lax-Friedrichs:

$$\begin{aligned} f^{LF}(a,b;n) &= \frac{1}{2}\left[n(f(a)+f(b)) - C(b-a)\right], \\ C &= \max\left\{|f'(s)| : \ \min(a,b) \leq s \leq \max(a,b)\right\}. \end{aligned}$$

Para dilucidar el funcionamiento de los flujos numéricos, pensemos un momento en el método de Gudonov para leyes de conservación escalares, esto es,

$$u_{n+1}^i = u_n^i - \lambda(f_{i+1/2}^* + f_{i-1/2}^*),$$

donde $u_n^i \approx u(x_i, t_n)$ y $f_{i+1/2}^*$ es el flujo numérico en $x_{i+1/2}$. Este último esta dado por $f_{i+1/2}^* = f(u^*)$ donde u^* se obtiene evaluando el problema de Riemann

$$u(x,t_n) = \begin{cases} u_n^i & \text{si } x < x_{i+1/2} \\ u_n^{i+1} & \text{si } x > x_{i+1/2} \end{cases}$$

en $(x_{i+1/2}, t_{n+1})$. Este método es conocido como *Riemann Solvers.*

En métodos numéricos iterativos es muy costoso y complicado evaluar u^* exactamente, por lo tanto es preferible una aproximación del problema de Riemann, las aproximaciones más sencillas son los flujos numéricos antes descritos. En [25] se estudian a fondo la construcción de las aproximaciones del problema de Riemann.

Al elegir el flujo numérico es posible establecer el esquema

$$\begin{aligned} \sum_{k=0}^{p} \frac{du_k^i(t)}{dt} \int_{I_i} N_k^i(x) N_l^i(x) dx &= \int_{I_i} f\left(\sum_{k=0}^{p} u_k^i(t) N_k^i(x)\right) \partial_x N_l^i(x) dx \\ &- \left[f\left(u_{i+1/2}^*\right) N_l^i(x_{i+1/2}) + f\left(u_{i-1/2}^*\right) N_l^i(x_{i-1/2})\right], \end{aligned}$$

para $l = 0, 1, \ldots, p$.

Sea $\mathbf{u}^i(t) = (u_0^i, \ldots, u_p^i)$, entonces podemos escribir el sistema de ecuaciones diferenciales ordinarias

$$\frac{d}{dt}\mathbf{u}^i(t) = \mathcal{L}(\mathbf{u}^i(t); u_0^{i+1}(t), u_p^{i-1}(t)). \tag{4.4}$$

Al analizar la segunda identidad en (4.1), se desprende que la condición inicial está dada por

$$\sum_{k=0}^{p} u_k^i(0) \int_{I_i} N_k^i(x) N_l^i(x) dx = \int_{I_i} u_0(x) N_l^i(x) dx. \tag{4.5}$$

Observemos que $u_0^{i+1}(t)$ y $u_p^{i-1}(t)$ son consideramos parámetros conocidos cuando aplicamos un método tipo Runge-Kutta a (4.4), (4.5).

En el siguiente teorema establecemos que si la solución numérica es convergente, entonces converge a cierto tipo de soluciones.

Teorema 4.1.1 *Sea f una función estrictamente convexa o cóncava en el problema de Cauchy* (4.1)*. Entonces para cualquier $p \geq 0$, si la solución numérica reconstruida mediante el método de Galerkin Discontinuo es convergente, entonces converge a la solución de entropía.*

Para una demostración de este teorema puede revisarse [3].

Para discretizar el sistema de ecuaciones diferenciales ordinarias (4.4) y las condiciones iniciales (4.5) en el tiempo usaremos un método tipo TVD Runge-Kutta. Procedemos de la siguiente forma: Sea $\{t^j\}_{j=0}^N$ una partición de $[0,T]$ y $\Delta t^j = t^{j+1} - t^j$, $j = 0, \ldots N-1$ nuestro algoritmo de marcha en el tiempo se describe como

- Sea $\mathbf{u}_0^i = \mathbf{u}^i(0)$
- Para $j = 0, \ldots, N-1$ calcular $\mathbf{u}_{j+1}^i$ a partir de $\mathbf{u}_j^i$ de la siguiente forma:

  1. $\mathbf{u}_{(0)}^i = \mathbf{u}_j^i$;
  2. para $l = 1, \ldots, p+1$ calcular las funciones intermedias

$$\mathbf{u}_{(l)}^i = \left\{ \sum_{k=0}^{l-1} \alpha_{lk} \mathbf{u}_{(k)}^i + \beta_{lk} \Delta t^j \mathcal{L}(\mathbf{u}_{(k)}^i; \theta) \right\};$$

  3. $\mathbf{u}_{j+1}^i = \mathbf{u}_{(k+1)}^i$.

Este método es sencillo de codificar ya que sólo es necesaria una subrutina que defina al operador $\mathcal{L}$. En la siguiente tabla se muestran algunos parámetros para el método Runge-Kutta.

Parámetros Método Runge-Kutta			
p	α_{lk}	β_{lk}	máx$\{\beta_{lk}/\alpha_{lk}\}$
2	$\begin{matrix} 1 & \\ \frac{1}{2} & \frac{1}{2} \end{matrix}$	$\begin{matrix} 1 & \\ 0 & \frac{1}{2} \end{matrix}$	1
3	$\begin{matrix} 1 & & \\ \frac{3}{4} & \frac{1}{4} & \\ \frac{1}{3} & 0 & \frac{2}{3} \end{matrix}$	$\begin{matrix} 1 & & \\ 0 & \frac{1}{4} & \\ 0 & 0 & \frac{2}{3} \end{matrix}$	1

Observemos que los valores para el parámetro α_{lk} mostrados en la tabla anterior son no negativos; esto no es coincidencia. De hecho, esta condición sobre α_{lk} asegura la propiedad

de estabilidad

$$|u_{j+1}^i| \leq |u_j^i|,$$

siempre que

$$|w| \leq |v|, \tag{4.6}$$

donde w se obtiene a partir de v con el método de Euler hacia adelante,

$$w = v + \delta \mathcal{L}(v), \tag{4.7}$$

para $|\delta|$ más pequeño que un δ_0 dado.

Por ejemplo, el método de Runge-Kutta de segundo orden mostrado en la tabla anterior se escribe como

$$\begin{aligned} u_{(1)}^i &= u_j^i + \Delta t \mathcal{L}(u_j^i), \\ w^i &= u_{(1)}^i + \Delta t \mathcal{L}(u_{(1)}^i), \\ u_{j+1}^i &= \frac{1}{2}\left(u_j^i + w^i\right). \end{aligned}$$

Ahora, si suponemos que las propiedades de estabilidad (4.6), (4.7) son satisfechas para

$$\delta_0 = |\Delta t \, \text{máx}\{\beta_{lk}/\alpha_{kl}\}| = \Delta t,$$

entonces,

$$|u_{(1)}^i| \leq |u_j^i|, \quad |w^i| \leq |u_{(1)}^i|,$$

y por lo tanto,

$$|u_{j+1}^i| \leq \frac{1}{2}\left(|u_j^i| + |w^i|\right) \leq |u_j^i|.$$

Observemos que podemos obtener este resultado debido a que los parámetros α son positivos. El ejemplo anterior, nos da una idea de como probar el siguiente resultado más general:

Teorema 4.1.2 (Estabilidad del Método de Runge-Kutta) *Supongamos que las propiedades de estabilidad* (4.6), (4.7) *para el método de Euler hacia adelante son satisfechas para*

$$\delta_0 = \min_{0 \leq j \leq N} |\Delta t^j \, \text{máx}\{\beta_{kl}/\alpha_{kl}\}|.$$

Más aún, supongamos que los coeficientes α_{kl} son no negativos y cumplen la propiedad

$$\sum_{l=0}^{k-1} \alpha_{kl} = 1, \quad k = 1, \ldots, p+1.$$

Entonces,

$$|u_j^i| \leq |u_0^i|, \quad \forall j \geq 0.$$

Esta propiedad de los métodos TVD-Runge-Kutta es crucial ya que nos permite obtener la estabilidad a partir de un método de Euler hacia adelante. Cuando se usan polinomios de grado p se requiere usar un método de Runge-Kutta de orden $p+1$. Para garantizar la estabilidad del método de Galerkin discontinuo, el enésimo paso temporal en el esquema Runge-Kutta se debe satisfacer

$$\Delta t^n \leq \frac{\Delta x}{\lambda_{\max}(2p+1)}, \tag{4.8}$$

donde $\lambda_{\max}$ es la velocidad característica máxima al tiempo t^n.

En [15] se estudian métodos de orden $s > 4$ que satisfacen la condición de no negatividad para los coeficientes α y β correspondientes; y se desarrollan técnicas de estabilización para sistemas de ecuaciones ordinarias que provienen de la discretización de ecuaciones diferenciales parciales.

Limitadores de Pendiente

Sea $\overline{u^i}$ el promedio de $u^i(x,t)$ en el conjunto I_i, esto es

$$\overline{u^i} = \fint_{I_i} u^i(x,t)dx.$$

Si $\varphi(x) = 1$ (aproximación mediante funciones constantes a trozos) en (4.2), entonces para la aproximación (4.3) podemos establecer

$$\int_{I_i} \partial_t u^i(x,t)dx + \left[f^*_{i+1/2} + f^*_{i-1/2}\right] = 0,$$

lo cual podemos, convenientemente reescribir como

$$\left[\overline{u^i}\right]_t + \frac{1}{\Delta_i}\left[f^*_{i+1/2} + f^*_{i-1/2}\right] = 0.$$

Esto demuestra que si w^i se obtiene mediante el método de Euler hacia adelante, esto es, $w^i = u^i + \delta \mathcal{L} u^i$, obtenemos

$$\frac{\overline{w^i} - \overline{u^i}}{\delta} + \frac{1}{\Delta_i}\left[f^*_{i+1/2} + f^*_{i-1/2}\right] = 0. \tag{4.9}$$

Cuando aproximamos la solución mediante funciones constantes a trozos, obtenemos un esquema monótono para valores suficientemente pequeños de $|\delta|$ y, como consecuencia, el esquema es de variación total decreciente (TVD por sus siglas en inglés, *Total Variation Diminishing*), esto es,

$$|\overline{w^i}|_{TV} \leq |\overline{u^i}|_{TV},$$

donde

$$|\overline{u^i}|_{TV} \equiv \sum_{i=1}^{M-1} |\overline{u^{i+1}} - \overline{u^i}|,$$

es la variación local de los promedios. Harten en [16] establece un resultado análogo para aproximaciones más generales, que nos dice cuando el esquema es TVD en promedios (TVDM).

Proposición 4.1.3 (Lema de Harten) *Si un esquema numérico con condiciones de frontera de soporte compacto o periódicas puede ser escrito en la forma*

$$\overline{w^i} = \overline{u^i} + C_{i+1/2}\left(\overline{u^{i+1}} - \overline{u^i}\right) - D_{i-1/2}\left(\overline{u^i} - \overline{u^{i-1}}\right) \tag{4.10}$$

con $C_{i+1/2}$ y $D_{i-1/2}$ funciones no lineales de $\overline{u^i}$, $\overline{u^{i\pm1}}$, $f^*_{i\pm1/2}$ que satisfacen

$$C_{i+1/2} \geq 0, \quad D_{i+1/2} \geq 0, \quad C_{i+1/2} + D_{i+1/2} \leq 1, \tag{4.11}$$

entonces es TVDM, es decir,

$$|\overline{w^i}|_{TV} \leq |\overline{u^i}|_{TV}.$$

Demostración: Observemos que

$$\begin{aligned}
\overline{w^{i+1}} - \overline{w^i} &= \left[1 - \left(C_{i+1/2} + D_{i+1/2}\right)\right]\left(\overline{u^{i+1}} - \overline{u^i}\right) + C_{i+3/2}\left(\overline{u^{i+2}} - \overline{u^{i+1}}\right) + D_{i-1/2}\left(\overline{u^i} - \overline{u^{i-1}}\right) \\
&= \left[1 - \left(C_{i+1/2} + D_{i+1/2}\right)\right]\left(\overline{u^{i+1}} - \overline{u^i}\right) + C_{j+1/2}\left(\overline{u^{j+1}} - \overline{u^j}\right) + D_{k+1/2}\left(\overline{u^{k+1}} - \overline{u^k}\right),
\end{aligned}$$

para $j = i+1$, $k = i-1$. De esta manera,

$$|\overline{w^{i+1}} - \overline{w^i}| \leq \left[1 - \left(C_{i+1/2} + D_{i+1/2}\right)\right]\left|\overline{u^{i+1}} - \overline{u^i}\right| + C_{j+1/2}\left|\overline{u^{j+1}} - \overline{u^j}\right| + D_{k+1/2}\left|\overline{u^{k+1}} - \overline{u^k}\right|.$$

Por lo tanto

$$\sum_i |\overline{w^{i+1}} - \overline{w^i}| \leq \sum_i \left[1 - \left(C_{i+1/2} + D_{i+1/2}\right) + C_{i+1/2} + D_{i+1/2}\right] \left|\overline{u^{i+1}} - \overline{u^i}\right|.$$

Así,

$$|\overline{w^i}|_{TV} \leq |\overline{u^i}|_{TV},$$

que es lo que se quería. □

Es fácil reescribir (4.9) en la forma (4.10). En efecto, sea

$$g_i^* = f^*(u_{i+1/2}^-, u_{i-1/2}^+, 1)$$

entonces

$$\begin{aligned} C_{i+1/2} &= -\frac{\delta}{\Delta_i}\left(\frac{f_{i+1/2}^* - g_i^*}{\overline{u^{i+1}} - \overline{u^i}}\right), \\ D_{i-1/2} &= \frac{\delta}{\Delta_i}\left(\frac{f_{i+1/2}^* + g_i^*}{\overline{u^i} - \overline{u^{i-1}}}\right). \end{aligned}$$

Así, los coeficientes $C_{i+1/2}$ y $D_{i-1/2}$ son positivos si y solo si se satisfacen las siguientes condiciones de signo

$$\begin{aligned} \text{sign}(u_{i+1/2}^+ - u_{i-1/2}^+) &= \text{sign}(\overline{u^{i+1}} - \overline{u^i}), \\ \text{sign}(u_{i+1/2}^- - u_{i-1/2}^-) &= \text{sign}(\overline{u^i} - \overline{u^{i-1}}), \end{aligned}$$

por la monotonicidad del flujo numérico f^*. Una vez que estas condiciones son satisfechas, la tercera condición en (4.11) se convierte en una restricción para el tamaño del parámetro δ.

Ya que la reconstrucción temporal tipo TVDRK que hemos usado en el método de Galerkin Discontinuo no produce soluciones que satisfacen las condiciones de signo anteriores, es necesario modificar el algoritmo, este proceso es conocido como limitadores de pendiente, y una de sus bondades es evitar oscilaciones espurias en la aproximación numérica.

A continuación construiremos un operador $\Lambda\Pi$ que satisface

- Mantiene la conservación de masa elemento a elemento. Si $v^i = \Lambda\Pi u^i$, entonces $\overline{v^i} = \overline{u^i}, \quad i = 1, \ldots, M.$

- Satisface las propiedades de signo. Esto asegura que si $v^i = \Lambda\Pi u$ y $w^i = v^i + \delta\mathcal{L}v^i$, entonces $|\overline{w^i}|_{TV} \leq |\overline{v^i}|_{TV}$.

La construcción del operador $\Lambda\Pi$ estará en términos de la función minmod,

$$\text{minmod}(a_1,\ldots,a_\nu) = \begin{cases} \alpha \min\limits_{1\le i\le \nu} |a_i| & \text{si } \alpha = \text{sign}(a_1) = \ldots = \text{sign}(a_\nu), \\ 0 & \text{Otro caso.} \end{cases}$$

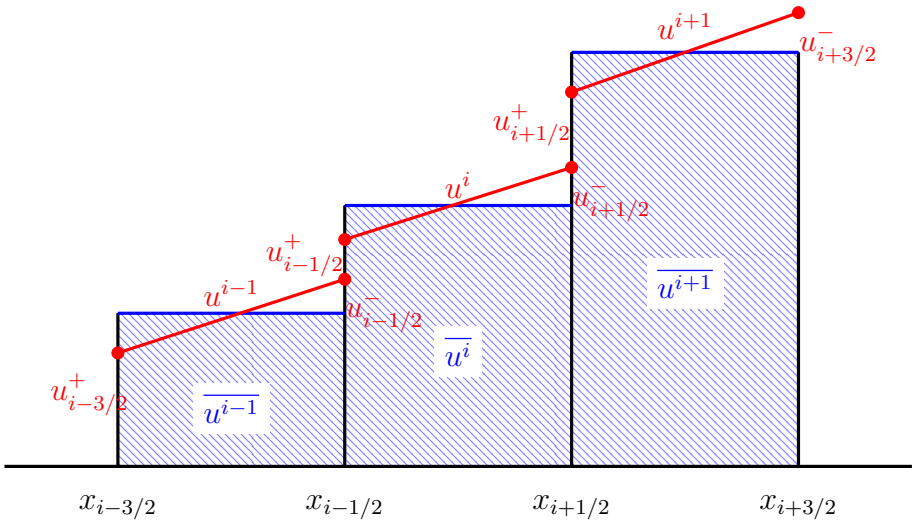


Figura 4.3: Reconstrucción de la solución u usando limitadores de pendiente.

Si usamos aproximaciones lineales a trozos, entonces

$$u^i = \overline{u^i} + (x - x_i)\partial_x u^i, \quad i = 1,\ldots,M,$$

de esta manera el operador $v^i = \Lambda\Pi u^i$ descrito como

$$v^i = \overline{u^i} + (x - x_i)\,\text{minmod}\left(\partial_x u^i, \frac{\overline{u^{i+1}} - \overline{u^i}}{\Delta_i}, \frac{\overline{u^i} - \overline{u^{i-1}}}{\Delta_i}\right)$$

satisface las propiedades antes descritas. Este limitador de pendiente es conocido como MUSCL (Monotone Upstream-centered Schemes for Conservation Laws) introducido por Van Leer en [40]. Observemos que la función minmod controla el crecimiento de la pendiente (oscilaciones) entre las interfaces $x_{i-1/2}$ y $x_{i+1/2}$, es decir, en los puntos de discontinuidad, ver gráfica 4.3.

El siguiente limitador de pendiente menos restrictivo también cumples las propiedades

$$v^i = \overline{u^i} + (x - x_i)\,\text{minmod}\left(\partial_x u^i, \frac{\overline{u^{i+1}} - \overline{u^i}}{\Delta_i/2}, \frac{\overline{u^i} - \overline{u^{i-1}}}{\Delta_i/2}\right).$$

Más aún, este limitador de pendiente lo podemos escribir como

$$\begin{aligned} v^-_{i+1/2} &= \overline{u^i} + \text{minmod}(u^-_{i+1/2} - \overline{u^i}, \overline{u^i} - \overline{u^{i-1}}, \overline{u^{i+1}} - \overline{u^i}), \\ v^-_{i-1/2} &= \overline{u^i} - \text{minmod}(\overline{u^i} - u^+_{i-1/2}, \overline{u^i} - \overline{u^{i-1}}, \overline{u^{i+1}} - \overline{u^i}). \end{aligned} \tag{4.12}$$

Denotaremos este limitador como $\Lambda\Pi^1$.

En el caso en que la aproximación $u^i(x,t)$ sea un polinomio de grado p, esto es

$$u^i(x,t) = \sum_{k=0}^{p} u_k^i(t) N_k^i(x),$$

donde $N_k^i(x)$ son los polinomios de Lagrange. Sea φ^i la parte lineal de u^i, entonces construimos $\Lambda\Pi^p$ de la siguiente manera

Algorithm 2: Construcción de $\Lambda\Pi^p$.

1 **for** $i = 1, \ldots, M$ **do**

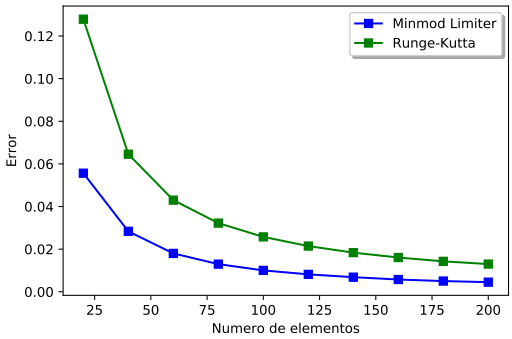
1. Calcular $v_{i+1/2}^-$ y $v_{i-1/2}^+$ usando (4.12).
2. Si $v_{i+1/2}^- = u_{i+1/2}^-$ y $v_{i-1/2}^+ = u_{i-1/2}^+$, entonces $v^i = u^i$.
3. Si no, $v^i = \Lambda\Pi^1 \varphi^i$.

En [21] Krivodonova describe limitadores de pendiente más generales para el método de Galerkin Discontinuo en una y dos dimensiones.

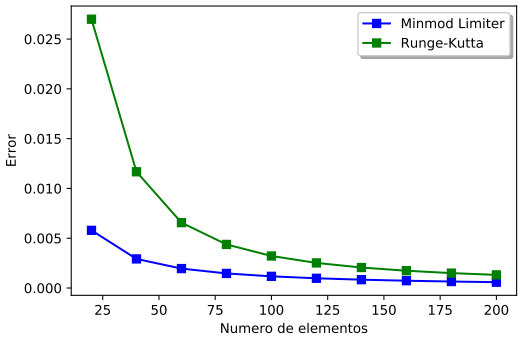
Ya que hemos descrito la construcción del operador $\Lambda\Pi$, a continuación estableceremos la discretización temporal para resolver numéricamente el problema (4.4). Sea $\{t^j\}_{j=0}^N$ una partición de $[0,T]$ y $\Delta t^j = t^{j+1} - t^j$, $j = 0, \ldots N-1$ nuestro algoritmo de marcha en el tiempo considerando limitadores de pendiente se describe como

- Sea $\mathbf{u}_0^i = \Lambda\Pi^p \mathbf{u}^i(0)$
- Para $j = 0, \ldots, N-1$ calcular $\mathbf{u}_{j+1}^i$ a partir de $\mathbf{u}_j^i$ de la siguiente forma:
  1. $\mathbf{u}_{(0)}^i = \mathbf{u}_j^i$;
  2. para $l = 1, \ldots, p+1$ calcular las funciones intermedias
  $$\mathbf{u}_{(l)}^i = \Lambda\Pi^p \left\{ \sum_{k=0}^{l-1} \alpha_{lk} \mathbf{u}_{(k)}^i + \beta_{lk} \Delta t^j \mathcal{L}(\mathbf{u}_{(k)}^i; \theta) \right\};$$
  3. $\mathbf{u}_{j+1}^i = \mathbf{u}_{(k+1)}^i$.

Este algoritmo describe completamente el método RKDG. Observemos que los limitadores de pendiente son calculados para cada función intermedia del método de Runge-Kutta. Este algoritmo en el paso temporal asegura que el esquema es TVDM.



Aproximación lineal, $p = 1$.



Aproximación cuadrática, $p = 2$.

Figura 4.4: Error en la norma $\mathbb{L}^2$ con y sin el uso de limitadores.

A continuación analizaremos el comportamiento de las soluciones del método de Galerkin Discontinuo usando limitadores de pendiente en el problema de Cauchy

$$\begin{aligned} u_t + f(u)_x &= 0 \\ u(x,0) &= u_0(x), \end{aligned}$$

con condiciones de frontera periódicas.

Ejemplo 4.1.1 *Consideremos la ecuación de advección con velocidad $a = 1$, es decir, $f(u) = u$ en el dominio $[-1,1] \times [0,2]$ y condición inicial*

$$u_0(x) = \sin(\pi x).$$

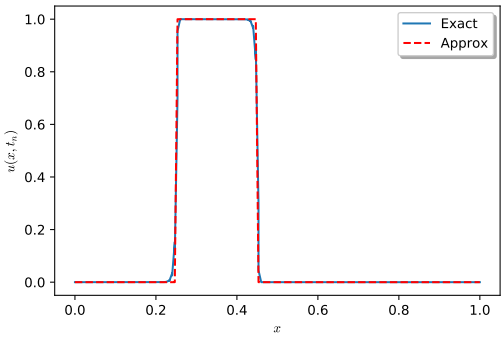
La solución exacta de este problema es suave y no requiere el uso de limitadores. En las gráficas 4.4 comparamos el error en la norma $\mathbb{L}^2$ con y sin limitadores de pendiente. Hemos usado distinto número de elementos espaciales con aproximaciones lineales y cuadráticas.

Ejemplo 4.1.2 *Volvamos a considerar la ecuación de advección con condición inicial*

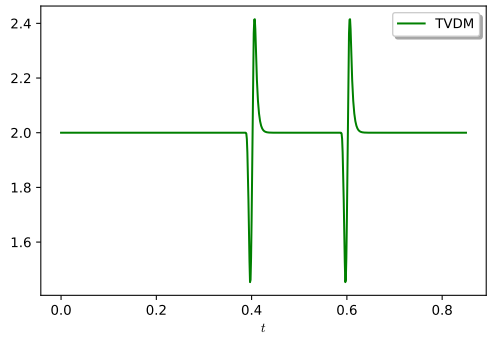
$$u_0(x) = \begin{cases} 1 & 0.4 < x < 0.6, \\ 0 & \textit{Otro caso}, \end{cases}$$

en el espacio $[0,1] \times [0,0.85]$. La solución analítica es un perfil que se mueve hacia la derecha, en la gráfica 4.5 mostramos una comparación entre la aproximación obtenida usando limitadores de pendiente con aproximaciones lineales con 150 elementos espaciales y la solución exacta; del lado derecho podemos observar la variación total en promedios.

Ejemplo 4.1.3 *Consideremos la ecuación de Burgers, es decir, $f(u) = \frac{1}{2}u^2$, en el domi-*



Reconstrucción numérica.



Variación Total en Promedios.

Figura 4.5: Método de Galerkin para la ecuación de advección con condición inicial discontinua.

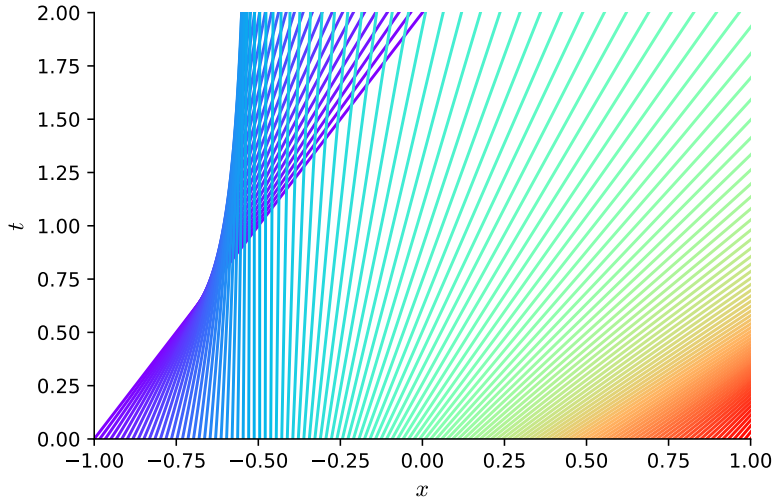


Figura 4.6: Rectas características para la ecuación de Burgers.

nio $[-1,1]\times[0,2]$ *y con condición inicial*

$$u_0(x)=\frac{1}{2}\left[1+\sin(\pi x)\right].$$

Observemos que el tiempo de rompimiento es $t=2/\pi$*, cuando se forma un* shock*. En la gráfica 4.6 podemos observar las rectas características. Hemos reconstruido la solución usando aproximaciones lineales y 100 elementos espaciales. En 4.7 mostramos los resultados con y sin limitadores de pendientes.*

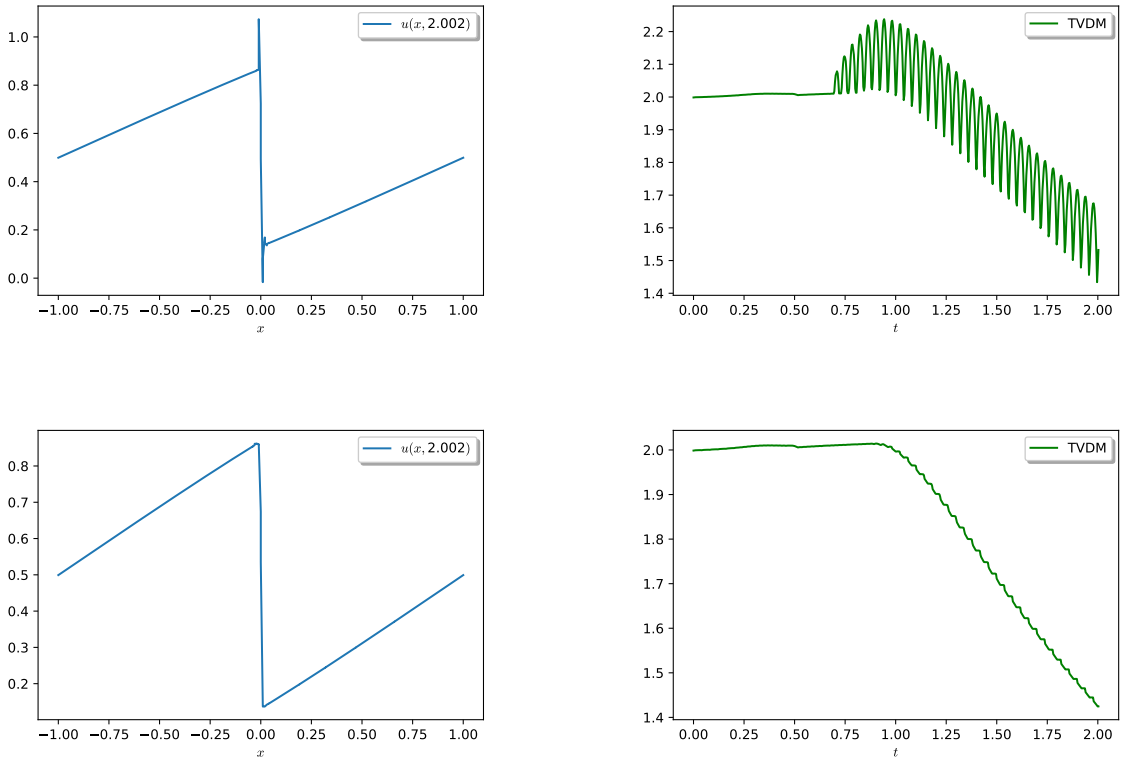


Figura 4.7: Reconstrucción numérica de la solución. En la parte superior no se usan limitadores y la inferior usamos el limitador minmod. Las gráficas del lado derecho corresponden a la variación total en promedios de la función $u(x,t)$.

Refinamiento y Convergencia

Usando el método de Galerkin discontinuo podemos encontrar una mejor aproximación numérica a la solución exacta a través de un proceso llamada refinamiento. Podemos considerar un número mayor de elementos I_i en el dominio, llamado h- refinamiento o aumentando la dimensión del espacio $\mathbb{V}(I_i)$, conocido como p-refinamiento, esto es, consideramos un orden local mayor de aproximación para obtener una mayor exactitud.

Para demostrar los efectos numéricos del refinamiento, consideremos la ley de conservación

$$\partial_t u + \partial_x u = 0, \quad (x,t) \in [0,1] \times (0,1),$$

con condición inicial

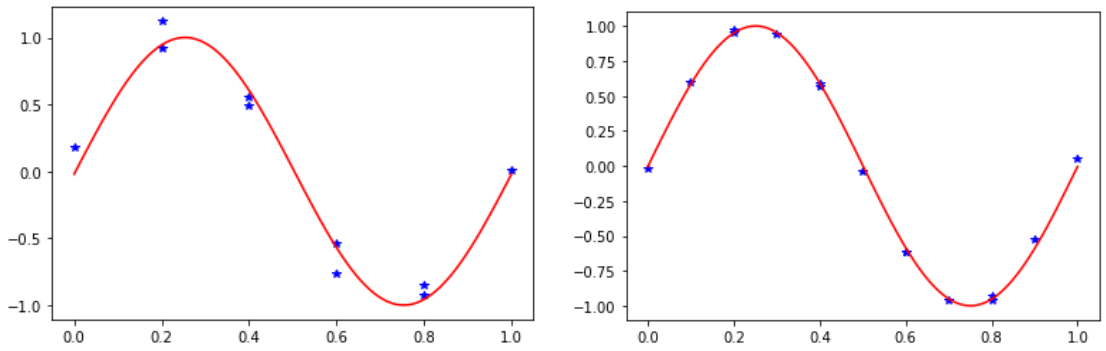
$$u_0(x) = \sin(2\pi x)$$

y condiciones de frontera periódicas. Podemos obtener la solución analítica de este pro-

blema a través del método de características:

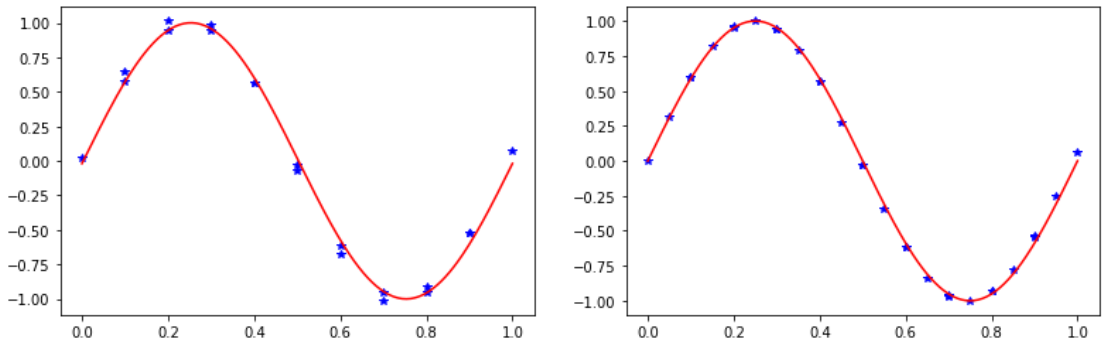
$$u(x,t) = \sin(2\pi(x-t)).$$

En las gráficas 5.7 y mostramos los resultados obtenidos mediante el método de Galerkin discontinuo, en rojo la solución analítica para un tiempo $t = 1$ y en azul la aproximación numérica. En la gráfica 4.10 mostramos la reconstrucción de la solución $u(x,t)$ en el espacio $[0,1] \times [0,1]$.



Aproximación lineal, $p = 1$. Aproximación cuadrática, $p = 2$.

Figura 4.8: p-refinamiento con 5 elementos.



Aproximación lineal, $p = 1$. Aproximación cuadrática, $p = 2$.

Figura 4.9: p-refinamiento con 10 elementos.

Para demostrar la eficacia de la combinación de los dos tipos de refinamientos, calculamos el error con la norma en $\mathbb{L}^2(\Omega)$

$$\left\|u - u^i\right\|_{\mathbb{L}^2(\Omega)} = \sum_{i=1}^{M} \sqrt{\int_{\Omega} (u - u^i)^2 dxdt}.$$

Los errores para diferentes niveles de (h,p)-refinamientos son mostrados en la tabla 4.1 con orden de convergencia. Mientras que el p-refinamiento parece ofrecer una mayor precisión sobre el $h-$refinamiento, pero es más costoso computacionalmente. Por otro lado,

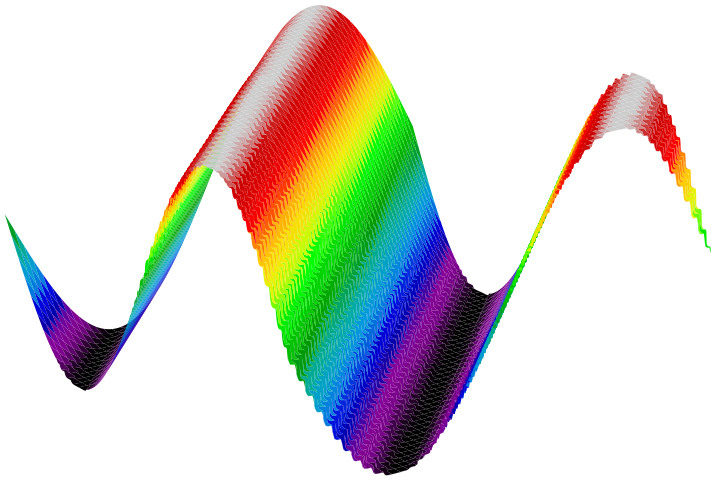


Figura 4.10: Reconstrucción de la función $u(x,t)$ mediante el método de Galerkin discontinuo.

el h-refinamiento demanda más memoria que el $p-$refinamiento. Combinando ambos tipos de refinamiento obtenemos una poderosa herramienta para incrementar la aproximación global.

	$p=1$		$p=2$		$p=3$		$p=4$	
M	Error	r	Error	r	Error	r	Error	r
10	3.06×10^{-2}		1.211×10^{-3}		4.650×10^{-5}		1.060×10^{-6}	
20	6.605×10^{-3}	2.238	1.513×10^{-4}	3.001	2.920×10^{-6}	3.993	3.337×10^{-8}	4.990
40	1.535×10^{-3}	2.084	1.891×10^{-5}	3.000	1.826×10^{-7}	4.000	1.054×10^{-9}	4.984
80	3.776×10^{-4}	2.023	2.364×10^{-6}	3.000	1.141×10^{-8}	4.000	3.325×10^{-11}	4.987

Cuadro 4.1: Error en la norma $\mathbb{L}^2$ y razón de convergencia.

4.2. Aplicación a las Ecuaciones de Saint-Venant

En esta sección describiremos a detalle el esquema para resolver las ecuaciones de Saint-Venant usando el método de Galerkin discontinuo y así como la codificación de este.

Recordemos que las ecuaciones de Saint-Venant cuando consideramos una batimetría no nula, las podemos escribir como

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t u + \partial_x\left(gh + \frac{1}{2}u^2\right) = -gB_x(x) \end{cases} \qquad x \in [a,b], \quad t \geq 0. \tag{4.13}$$

Para hallar una solución numérica de (4.13) aplicaremos el método de Galerkin discontinuo descrito en la sección 4.1.

Al considerar la primera ecuación, para $l = 0, \ldots, p$, podemos escribir

$$\begin{aligned} \sum_{k=0}^{p} \frac{d}{dt} h_k^i(t) \int_{I_i} N_k^i(x) N_l^i(x) dx &= \int_{I_i} \left[\sum_{k=0}^{p} u_k^i(t) N_k^i(x)\right] \left[\sum_{j=0}^{p} h_j^i(t) N_j^i(x)\right] \frac{d}{dx} N_l^i(x) dx \\ &- \left[\alpha_{i+1/2}^* N_l^i(x_{i+1/2}) + \alpha_{i-1/2}^* N_l^i(x_{i-1/2})\right], \end{aligned}$$

o equivalentemente

$$\begin{aligned} \sum_{k=0}^{p} \frac{d}{dt} h_k^i(t) \left\langle N_k^i(x), N_l^i(x)\right\rangle_{\mathbb{L}^2(I_i)} &= \left\langle h^i(x,t) u^i(x,t), \frac{dN_l^i(x)}{dx}\right\rangle_{\mathbb{L}^2(I_i)} \\ &- \left[\alpha_{i+1/2}^* N_l^i(x_{i+1/2}) + \alpha_{i-1/2}^* N_l^i(x_{i-1/2})\right]. \end{aligned}$$

De manera similar, para la segunda ecuación el esquema es descrito como

$$\begin{aligned} \sum_{k=0}^{p} \frac{d}{dt} u_k^i(t) \int_{I_i} N_k^i(x) N_l^i(x) dx &= \int_{I_i} \left(g \sum_{k=0}^{p} h_k^i(t) N_k^i(x) + \frac{1}{2} \left\{ \sum_{k=0}^{p} u_k^i(t) N_k^i(x) \right\}^2 \right) \frac{d}{dx} N_l^i(x) dx \\ &- \int_{I_i} g B_x(x) N_l^i(x) dx \\ &- \left[\beta_{i+1/2}^* N_l^i(x_{i+1/2}) + \beta_{i-1/2}^* N_l^i(x_{i-1/2})\right], \end{aligned}$$

que podemos escribir en la forma

$$\begin{aligned} \sum_{k=0}^{p} \frac{d}{dt} u_k^i(t) \left\langle N_k^i(x), N_l^i(x)\right\rangle_{\mathbb{L}^2(I_i)} &= \left\langle g h^i(x,t) + \frac{u^i(x,t)^2}{2}, \frac{dN_l^i(x)}{dx}\right\rangle_{\mathbb{L}^2(I_i)} \\ &- \left\langle g B_x(x), N_l^i(x)\right\rangle_{\mathbb{L}^2(I_i)} - \left[\beta_{i+1/2}^* N_l^i(x_{i+1/2}) + \beta_{i-1/2}^* N_l^i(x_{i-1/2})\right], \end{aligned}$$

donde $F^* = (\alpha^*, \beta^*)$ es el flujo numérico escogido para la función $F = (hu, gh + \frac{1}{2}u^2)$.

Observemos que hemos obtenido un sistema de ecuaciones diferenciales ordinarias de $(2p+2) \times (2p+2)$ que resolveremos numéricamente usando un método de Runge-Kutta con orden $p+1$.

Codificación

En esta sección describiremos de manera breve la codificación del esquema numérico de Galerkin discontinuo para las ecuaciones de aguas someras.

Sea I_i la celda de estudio. Definamos $\mathbf{\Phi^i(x)}, \mathbf{d\Phi^i(x)} \in \mathbb{R}^{p+1}$ vectores de funciones base y sus derivadas evaluadas en x, respectivamente, para la celda I_i, es decir,

$$\begin{aligned} \mathbf{\Phi}^i(x) &:= [N_0^i(x), N_1^i(x), \ldots, N_p^i(x)]^t, \\ \mathbf{d\Phi}^i(x) &:= [dN_0^i(x), dN_1^i(x), \ldots, dN_p^i(x)]^t. \end{aligned}$$

Más aún, sea $M^i \in \mathbb{R}^{(p+1)\times(p+1)}$ la matriz de masa

$$\{M^i\}_{kl} := \int_{I_i} N_k^i(x) N_l^i(x) dx.$$

Calcularemos cada entrada de la matriz anterior usando cuadraturas de Gauss, en [33] se estudian ampliamente.

Para codificar el esquema de Galerkin discontinuo necesitamos definir los vectores $\mathbf{h}^i, \mathbf{u}^i \in \mathbb{R}^{p+1}$ de la siguiente forma:

$$\mathbf{h}^i(t) := [h_0^i(t), h_1^i(t), \ldots, h_p^i(t)]^t \quad \mathbf{u}^i(t) := [u_0^i(t), u_1^i(t), \ldots, u_p^i(t)]^t.$$

Sean $F_h, F_\alpha \in \mathbb{R}^{p+1}$ los vectores

$$\{F_h\}_l = \int_{I_i} \left(\mathbf{u^i} \cdot \mathbf{\Phi^i(x)}\right) \left(\mathbf{h^i} \cdot \mathbf{\Phi^i(x)}\right) \mathbf{d\Phi^i(x)}_l dx,$$

$$\{F_\alpha\}_l = \alpha^*_{i+1/2} N_l^i(x_{i+1/2}) + \alpha^*_{i-1/2} N_l^i(x_{i-1/2}).$$

Ya que las funciones base tienen la propiedad $N_k^i(x_l^i) = \delta_{kl}$, entonces

$$F_\alpha = [\alpha^*_{i-1/2}, 0, \ldots, 0, \alpha^*_{i+1/2}]^t.$$

Finalmente, podemos escribir la codificación del esquema de Galerkin discontinuo para la ecuación

$$\partial_t h + \partial_x (hu) = 0,$$

como

$$M^i \frac{d\mathbf{h}^i(t)}{dt} = F_h - F_\alpha. \tag{4.14}$$

De manera similar, para

$$\partial_t u + \partial_x \left(gh + \frac{1}{2}u^2 \right) = -gB_x(x)$$

el esquema toma la forma

$$M^i \frac{d\mathbf{u}^i(t)}{dt} = F_u - F_B - F_\beta, \tag{4.15}$$

donde

$$\begin{aligned}
\{F_u\}_l &= \int_{I_i} \left(g\mathbf{h^i} \cdot \mathbf{\Phi^i}(\mathbf{x}) + \frac{\mathbf{1}}{\mathbf{2}} \left\{ \mathbf{u^i} \cdot \mathbf{\Phi^i}(\mathbf{x}) \right\}^{\mathbf{2}} \right) \mathbf{d\Phi^i}(\mathbf{x})_l dx, \\
\{F_B\}_l &= g \int_{I_i} B_x(x) \mathbf{d\Phi^i}(\mathbf{x})_l dx, \\
\{F_\beta\}_l &= [\beta^*_{i-1/2}, 0, \ldots, 0, \beta^*_{i+1/2}]^t.
\end{aligned}$$

Como la aproximación de Galerkin discontinuo en el intervalo I_i para $h(x,t)$ es

$$h^i(x,t) = \sum_{k=0}^{p} h^i_k(t) N^i_k(x),$$

entonces podemos inferir la condición inicial multiplicando por una función prueba e integrando sobre I_i

$$\sum_{k=0}^{p} h^i_k(0) \int_{I_i} N^i_k(x) N^i_l(x) dx = \int_{I_i} h_0(x) N^i_l(x) dx. \tag{4.16}$$

De esta manera obtenemos el sistema lineal de ecuaciones

$$M^i \mathbf{h}^i(0) = \left\{ \int_{I_i} h_0(x) N^i_l(x) dx \right\}_{l=0}^{p}.$$

De manera similar, podemos obtener $\mathbf{u}^i(0)$. Sea $\mathbf{w}^i \in \mathbb{R}^{2p+2}$ definido como

$$\mathbf{w}^i(t) = [\mathbf{h}^i(t), \mathbf{u}^i(t)]^t.$$

De esta manera, cuando usamos (4.14) y (4.15), podemos escribir el problema de valores iniciales

$$\frac{d\mathbf{w}^i(t)}{dt} = \mathcal{L}(\mathbf{w}^i, \theta_i), \quad \mathbf{w}^i(0) = [\mathbf{h}^i(0), \mathbf{u}^i(0)]^t,$$

donde θ_i son los parámetros necesarios para calcular F_α y F_β.

A continuación, a través de pseudo código describiremos la codificación del método de

Galerkin discontinuo.

Algorithm 3: Pseudo código de Galerkin Discontinuo con un esquema temporal Runge-Kutta

Input: p, s, x_L, x_R, m, $B(x)$, t_f
Output: U, $H \in \mathbb{R}^{(p+1)\times m\times(n+1)}$

$X = linspace(x_L, x_R, m+1)$
$M = inv(M_{\text{mass}}(p))$
`/* Establece condiciones iniciales para` h `y` u `*/`
$h0j = W_0(p, h_0, M, X)$
$u0j = W_0(p, u_0, M, X)$
`/* Cálculo de las constantes para el flujo de Lax-Friedrichs */`
$C, dx, dt = CFL(p, X, h0j, u0j)$
$n = \lceil \frac{t_f}{dt} \rceil + 1$
`/* Recorrido temporal */`
for $k = 1, \ldots, n+1$ **do**
  `/* Recorrido espacial */`
  for $i = 1, \ldots, m$ **do**
    $a_i = 0.5 \cdot (X[i+1] - X[i])$
    $M = (1/a_i)M^i$
    $h_i = H[:, i, k-1]$
    $u_i = U[:, i, k-1]$
    `/* Parámetros necesarios para construir los flujos */`
    $\theta_i = [H[p, i-1, k-1], U[p, i-1, k-1], H[0, i+1, k-1], U[0, i+1, k-1]]$
    `/* Resolver el problema de valores iniciales usando un método Runge-Kutta tipo TVD de grado` s `*/`
    $H[:, i, k], U[:, i, k] = \text{TVDRK}(p, s, Mi, dt, hi, ui, \theta_i, C, \mathcal{L})$

A continuación, explicaremos algunas ideas sobre las funciones en el pseudo código anterior.

La función $W_0(p, h_0, M, X)$ evalúa la expresión (4.16) para todo elemento I_i en la partición espacial X.

La codificación de $C, dx, dt = CFL(p, X, h0j, u0j)$ calcula la constante C para usar el flujo de Lax-Friedrichs global, dx el máximo de las longitudes de los elementos y el paso temporal dt mediante la condición de Courant-Friedrichs-Levy (4.8), donde

$$\lambda_{\max} = u + \sqrt{gh}$$

es calculado usando las condiciones iniciales $h_0(x)$ y $u_0(x)$.

Ya que la matriz de masa M^i para el elemento I_i es un matriz constante M multiplicada por el jacobiano, esto es

$$M^i = a_i M,$$

entonces basta calcular la matriz M (o su inversa) una sola vez, y en cada elemento multiplicarla por el jacobiano. Este proceso se observa en las líneas 2 y 10 del pseudo código.

Las condiciones de frontera son codificadas en la linea 13, dependiendo en que elemento nos encontremos debemos modificar el conjunto de parámetros θ_i, ya que tenemos condiciones de frontera en $x = a$, entonces

$$\theta_0 = [h_a((k-1)dt), u_a((k-1)dt), H[0,1,k-1], U[0,1,k-1]].$$

4.3. Aplicación al Sistema Adjunto

En esta sección describiremos el método de Galerkin discontinuo para resolver numéricamente el sistema (3.1).

Sean α y β definidas como

$$\alpha(x,t) := \lambda(x, T-t) \quad \beta(x,t) := \mu(x, T-t).$$

Consideremos las funciones f y w

$$f(x,t) := h(x, T-t), \quad w(x,t) := u(x, T-t).$$

De esta manera, podemos escribir el sistema adjunto como

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}_t - \begin{pmatrix} w & g \\ f & w \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_x = \begin{pmatrix} 0 \\ -\mathcal{M}^*(\mathcal{M}w - \hat{w}) \end{pmatrix}, \tag{4.17}$$

con condiciones iniciales y de frontera

$$\begin{aligned} \alpha(x,0) = 0, \quad \beta(x,0) = 0, \\ \alpha(b,t) = 0, \quad \beta(b,t) = 0. \end{aligned} \tag{4.18}$$

Al multiplicar cada ecuación del sistema (4.17) por una función prueba e integrar sobre una celda de control I_i, obtenemos

$$\begin{aligned} \int_{I_i} \alpha_t \varphi dx &= \varphi(w\alpha + g\beta)\Big|_{\partial I_i} - \int_{I_i} (w\alpha + g\beta)\frac{d\varphi}{dx}dx - \int_{I_i} \alpha\varphi\partial_x w dx, \\ \int_{I_i} \beta_t \varphi dx &= \varphi\left(f\alpha + w\beta\right)\Big|_{\partial I_i} - \int_{I_i} \left(f\alpha + w\beta\right)\frac{d\varphi}{dx} - \int_{I_i} \left(\alpha\partial_x f + \beta\partial_x w\right)\varphi dx \\ &- \int_{I_i} \mathcal{M}^*(\mathcal{M}w - \hat{w})\varphi dx. \end{aligned}$$

Si suponemos que

$$\alpha(x,t)\Big|_{I_i} = \sum_{k=0}^{p} \alpha_k^i(t) N_k^i(x) \quad \text{y} \quad \beta(x,t)\Big|_{I_i} = \sum_{k=0}^{p} \beta_k^i(t) N_k^i(x).$$

Observemos que f y w tienen expresiones similares, es decir,

$$f(x,t)\Big|_{I_i} = \sum_{k=0}^{p} f_k^i(t) N_k^i(x), \quad w(x,t)\Big|_{I_i} = \sum_{k=0}^{p} w_k^i(t) N_k^i(x).$$

Más aún, al reemplazar φ por $N_l^i(x)$ para algún $l \in \{0,1,\ldots,p\}$, obtenemos los esquemas

$$\begin{aligned}
\sum_{k=0}^{p} \frac{d\alpha_k^i(t)}{dt} \left\langle N_l^i(x), N_k^i(x) \right\rangle_{\mathbb{L}^2(I_i)} &= \left[N_l^i(x_{i+1/2}) a_{i+1/2}^* + N_l^i(x_{i-1/2}) a_{i-1/2}^* \right] \\
&- \left\langle \alpha^i(x,t) + g\beta^i(x,t), \frac{N_l^i(x)}{dx} \right\rangle_{\mathbb{L}^2(I_i)} \\
&- \left\langle \alpha^i(x,t) \partial_x w^i(x,t), N_l^i(x) \right\rangle_{\mathbb{L}^2(I_i)},
\end{aligned}$$

$$\begin{aligned}
\sum_{k=0}^{p} \frac{d\beta_k^i(t)}{dt} \left\langle N_l^i(x), N_k^i(x) \right\rangle_{\mathbb{L}^2(I_i)} &= \left[N_l^i(x_{i+1/2}) b_{i+1/2}^* + N_l^i(x_{i-1/2}) b_{i-1/2}^* \right] \\
&- \left\langle f^i(x,t)\alpha^i(x,t) + w^i(x,t)\beta^i(x,t), \frac{dN_l^i(x)}{dx} \right\rangle_{\mathbb{L}^2(I_i)} \\
&- \left\langle \alpha^i(x,t)\partial_x f^i(x,t) + \beta^i(x,t)\partial_x w^i(x,t), N_l^i(x) \right\rangle_{\mathbb{L}^2(I_i)} \\
&- \int_{I_i} M^*(\mathcal{M}w - \hat{w}) N_l^i(x) dx,
\end{aligned}$$

donde $F^* = (a^*, b^*)$ es el flujo numérico que escogemos para $F = (w\alpha + g\beta, f\alpha + \beta w)$. Observemos que a^* y b^* dependen de los valores extremos de $f^i(x,t)$ y $w^i(x,t)$.

En las siguientes líneas describiremos como calcular

$$\int_{I_i} \mathcal{M}^*(\mathcal{M}w - \hat{w}) N_l^i(x) dx.$$

Si consideramos los puntos de observación los mismos para los cuales construimos la solución numérica de las ecuaciones de aguas someras, h, u. Entonces, podemos escoger $x_j \in \{x_0^i, \ldots, x_p^i\}$ (partición de I_i). Sea $E_t = [t - \Delta t/2, t + \Delta t/2]$. De esta manera, al usar

lo expuesto en la sección 3.3, tenemos que

$$\begin{aligned}
\int_{I_i} \mathcal{M}^*(\mathcal{M}w - \hat{w})N_l^i(x)dx &\approx \frac{1}{\Delta t}\sum_{j,n}\left[w(x_j,t_n;B) - \hat{w}_{jn}\right]\left[N_l^i(x_j)\chi_{I_i}(x_j)\chi_{E_t}(t_n)\right] \\
&= \frac{1}{\Delta t}\sum_{j=0}^{p}\sum_{n}\left[w(x_j^i,t_n;B) - \hat{w}_{jn}\right]\left[N_l^i(x_j^i)\chi_{I_i}(x_j^i)\chi_{E_t}(t_n)\right] \\
&= \frac{1}{\Delta t}\sum_{j=0}^{p}\sum_{n}\left[w(x_j^i,t_n;B) - \hat{w}_{jn}\right]\left[\delta_{lj}\chi_{E_t}(t_n)\right] \\
&= \frac{1}{\Delta t}\sum_{n}\left[w(x_l^i,t_n;B) - \hat{w}_{ln}\right]\chi_{E_t}(t_n) \\
&= \frac{1}{\Delta t}\left[w(x_l^i,t_k;B) - \hat{w}_{lk}\right] \\
&= \frac{1}{\Delta t}\left[\sum_{k=0}^{p} w_k^i(t_k)N_k^i(x_l^i) - \hat{w}_{lk}\right] \\
&= \frac{1}{\Delta t}\left[w_l^i(t_k) - \hat{w}_{lk}\right],
\end{aligned}$$

donde t_k es el único punto que cumple $t_k \in E_t$.

A continuación esbozaremos una manera computacional de calcular $\nabla\mathcal{J}(B)$.

Recordemos que

$$\mu(x,t) = \bigoplus_{i=1}^{M}\mu^i(x,t), \quad \mu^i(x,t) = \sum_{k=0}^{p}\mu_k^i(t)N_k^i(x).$$

Por lo tanto,

$$\begin{aligned}
\nabla\mathcal{J}(B)\Big|_{I_i} &= -g\int_0^T \mu_x^i(x,t)dt \\
&= -g\int_0^T \frac{\partial}{\partial x}\left\{\sum_{k=0}^{p}\mu_k^i(t)N_k^i(x)\right\}dt \\
&= -g\sum_{k=0}^{p}\frac{dN_k^i(x)}{dx}\int_0^T \mu_k^i(t)dt
\end{aligned}$$

Cuando consideramos la malla temporal $\{t_j\}_{j=0}^N$, con $t_0 = 0$, $t_N = T$, tenemos que

$$\mu_k^i(t_j) \approx \mu_{k,j}^i, \quad t \in \mathcal{B}_{\frac{\Delta t}{2}}(t_j)$$

donde $\mu_{k,j}^i$ es la aproximación numérica obtenida del método Galerkin discontinuo. Al

construir bandas de tamaño Δt alrededor de los puntos $\{t_j\}$, podemos hacer la siguiente aproximación

$$\int_0^T \mu_k^i(t)dt \approx \sum_{j=-1}^{\frac{2N-3}{2}} \int_{\frac{2j+1}{2}\Delta t}^{\frac{2j+3}{2}\Delta t} \mu_k^i(t)dt \approx \Delta t \sum_{j=0}^{N} \mu_{k,j}^i.$$

De esta manera,

$$\nabla \mathcal{J}(B)\Big|_{I_i} \approx -g\Delta t \sum_{k=0}^{p} \sum_{j=0}^{N} \frac{dN_k^i(x)}{dx} \mu_{k,j}^i.$$

Capítulo 5

Resultados Numéricos

5.1. Aguas Someras

En esta sección consideraremos el modelo de las ecuaciones de Saint-Venant

$$\begin{cases} \partial_t h + \partial_x(hu) = 0 \\ \partial_t u + \partial_x\left(gh + \frac{1}{2}u^2\right) = -gB_x(x) \end{cases} \qquad x \in [a,b], \quad t \geq 0.$$

En cada ejemplo consideramos condiciones iniciales y de frontera adecuadas.

La metodología para resolver el problema inverso sera la siguiente:

1. Resolver las ecuaciones de Saint-Venant
2. Creamos datos sintéticos de observación de la siguiente manera

$$h_{\text{obs}} = h(1 + \mathcal{N}(0, \sigma^2)),$$

   donde $\mathcal{N}$ es la distribución normal.

3. Resolveremos las ecuaciones adjuntas y calcularemos numéricamente el gradiente del funcional $\mathcal{J}$.
4. Implementamos el método de optimización BFGS para reconstruir la batimetría, inicializando el algoritmo con una aproximación $B_0(x)$.

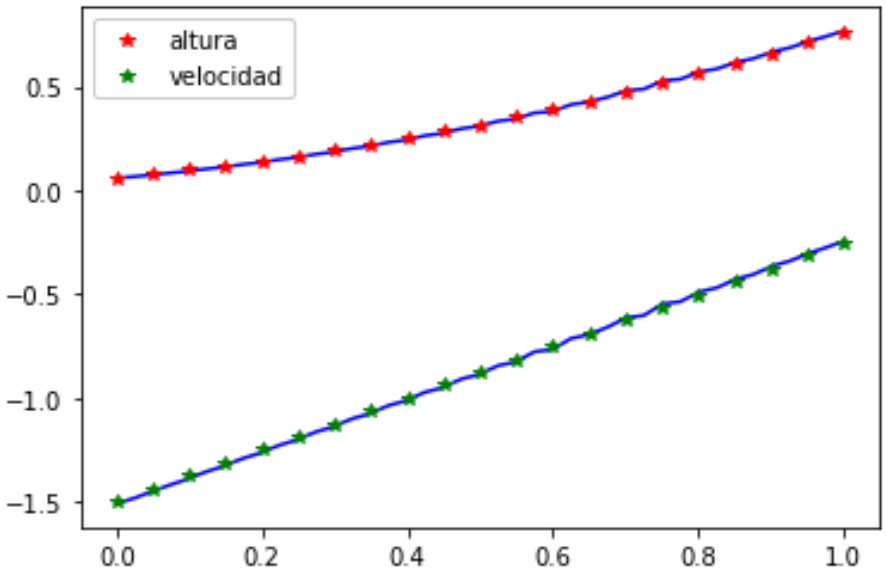


Figura 5.1: Aproximación numérica obtenida a través del método de Galerkin discontinuo al tiempo $t = 1.0$. La linea solida azul representa la solución analítica para la velocidad y la altura, mientras que los asteriscos es la aproximación.

5.2. Solución Analítica

Como un primer ejemplo, consideremos una batimetría plana, es decir, $B \equiv 0$, para este caso, se tiene una solución analítica de la forma

$$h(x,t) = \xi^2, \quad u(x,t) = 2\sqrt{g}\xi - 2\sqrt{gH},$$

donde

$$\xi(x,t) = \frac{x + 2\sqrt{gH}t}{1 + 3\sqrt{g}t},$$

y H es un estado estacionario de $h(x,t)$. Para este ejemplo hemos considerado una malla espacio-temporal contenida en $[0,1] \times [0,1]$. Ese problema es tratado en [18].

Este ejemplo nos proporciona una herramienta para estudiar la convergencia del método de Galerkin discontinuo para las ecuaciones de Saint-Venant.

Recordemos que estamos considerando el producto interno

$$\langle f, g \rangle_{\mathbb{L}^2(\Omega\times(0,T))} = \frac{1}{(b-a)T} \int_0^T \int_a^b f(x,t)g(x,t)dxdt,$$

entonces podemos calcular el error como

$$E_u := \sqrt{\Delta x \Delta t \sum_{j=1}^{M} \sum_{n=1}^{N} (u(x_j, t_n) - u^*(x_j, t_n))},$$

donde u^* es la aproximación numérica que obtuvimos a través del método de Galerkin discontinuo. De manera análoga podemos establecer el error E_h para la aproximación de la altura h. En la table 5.1 presentamos los errores obtenidos para distinto número de elementos espaciales y diferentes grados de aproximación.

	m_e	E_h	E_u
$p = 1$	20	$8.50E-02$	$1.22E-01$
	40	$4.29E-02$	$6.27E-02$
	80	$2.17E-02$	$3.19E-02$
	160	$1.08E-02$	$1.60E-02$
$p = 2$	20	$6.77E-03$	$4.30E-03$
	40	$1.78E-03$	$1.07E-03$
	80	$4.55E-04$	$2.74E-04$
	160	$1.14E-04$	$6.84E-04$
$p = 3$	20	$8.84E-05$	$1.13E-04$
	40	$1.12E-05$	$1.42E-05$
	80	$1.44E-06$	$1.80E-06$
	160	$1.72E-07$	$2.20E-07$

Cuadro 5.1: Errores E_h y E_u con m_e elementos espaciales usando polinomios de grado p.

5.3. Tope

Este es un ejemplo clásico en la literatura, se puede consultar, por ejemplo en [13].

Consideramos el espacio $[0, 25]$ con la batimetría

$$B(x) = \begin{cases} 0.2 - 0.05(x-10)^2 & 8 \leq x \leq 12 \\ 0.0 & \text{otro caso} \end{cases}, \tag{5.1}$$

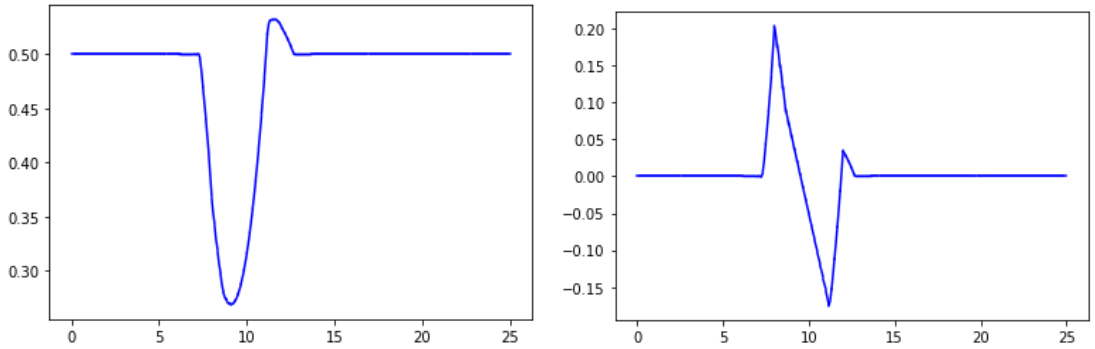
con condiciones iniciales

$$h_0(x) = 0.5 - B(x), \quad u_0(x) = 0.0,$$

y condiciones de frontera

$$h(0, t) = 0.5, \quad u(0, t) = 0.0.$$

En las gráficas 5.2 mostramos la aproximación numérica obtenida mediante el método de Galerkin discontinuo.



Aproximación numérica de la altura, h. Aproximación numérica de la velocidad, u.

Figura 5.2: Aproximación numérica mediante el método de Galerkin discontinuo para las ecuaciones de aguas someras al tiempo $t = 1.0$ y batimetría descrita en (5.1).

En la gráfica 5.3 podemos observar la función $\eta(x,t) = h(x,t) + B(x)$ para distintos tiempos, y la batimetría.

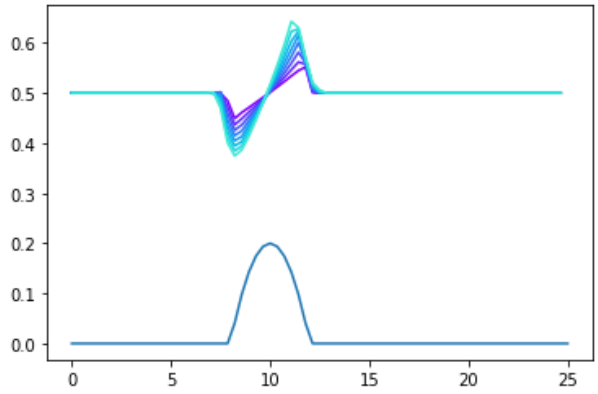


Figura 5.3: Gráfica de $\eta(x,t)$, $t = 0.3s, 0.4s, \ldots, 0.9s$.

En este caso hemos considerado 70 elementos espaciales, un tiempo final $t_f = 1.0$ y hemos inicializado el algoritmo de optimización con una aproximación inicial $B_0(x) = 0.8B(x)$. Mostramos los resultados obtenidos en la gráfica 5.4.

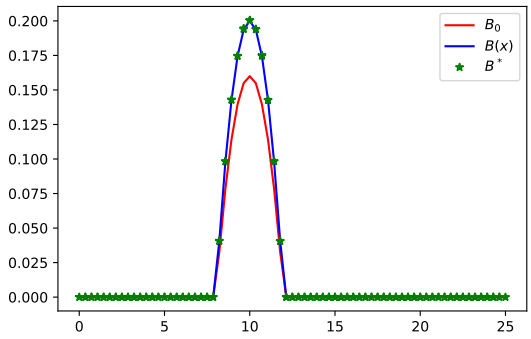


Figura 5.4: Reconstrucción de la batimetría. B_0 es la aproximación inicial, B^* es el resultado del algoritmo de optimización y $B(x)$ es la batimetría real.

5.4. Mareas Oceánicas

El primer ejemplo que analizaremos fue propuesto por Bermúdez y Vázquez [4] para estudiar las mareas en zonas costeras.

Sea

$$H(x) = 50.5 - \frac{40x}{L} + 10\sin\left\{\pi\left(\frac{4x}{L} + 0.5\right)\right\},$$

con $L = 648000$, e impongamos las condiciones iniciales

$$h(x,0) = H(x) \quad \text{y} \quad u(x,0) = 0, \quad x \in [0, L],$$

con condiciones de frontera

$$h(0,t) = \begin{cases} 64.5 + 4\sin\left\{\pi\left(\frac{4t}{86400} - 0.5\right)\right\} & t \leq 43200 \\ 60.5 & t > 43200 \end{cases},$$

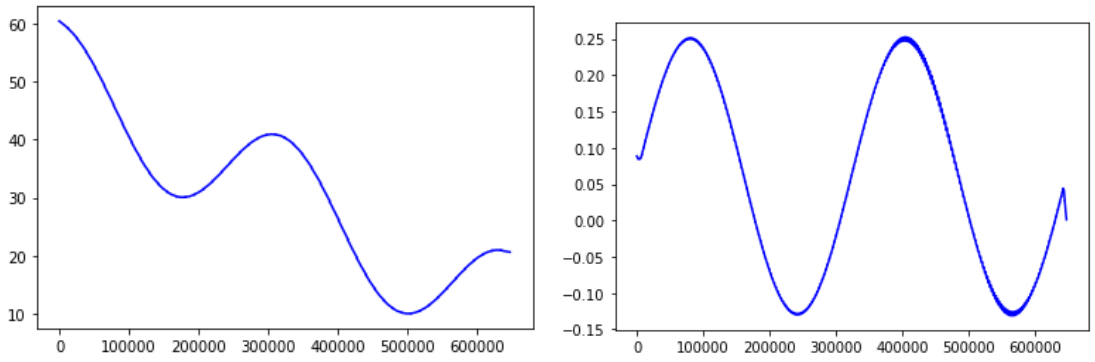
que representa una ola entrante y $u(L,t) = 0$. Para este problema, hemos considerado la batimetría

$$B(x) = 60.5 - H(x).$$

Este problema modela la propagación de mareas oceánicas. En este caso, $h(0,t)$ representa una marea de 4 metros de amplitud, podemos observar que al tiempo $t = 21600$ segundos, la ola alcanza su altura máxima de 8 metros, mientras que al tiempo 43200 segundos la ola desaparece.

En las gráficas 5.5 podemos observar la aproximación numérica para h y u usando el método de Galerkin discontinuo con una aproximación cuadrática y un tiempo final

$t_f = 1000s$.



Reconstrucción numérica de la altura, h. Reconstrucción numérica de la velocidad, u.

Figura 5.5: Reconstrucción numérica mediante el método de Galerkin discontinuo en el tiempo $t_f = 1000s$.

En la gráfica 5.6 podemos observar la función $\eta(x,t) = B(x) + h(x,t)$ para distintos tiempos.

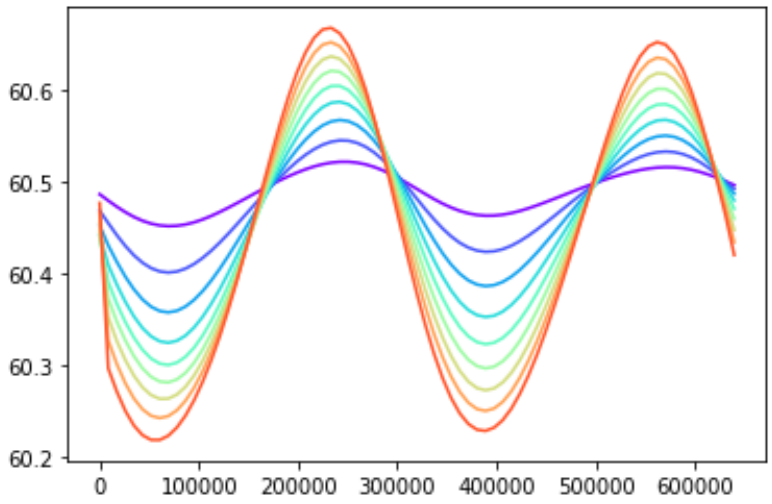
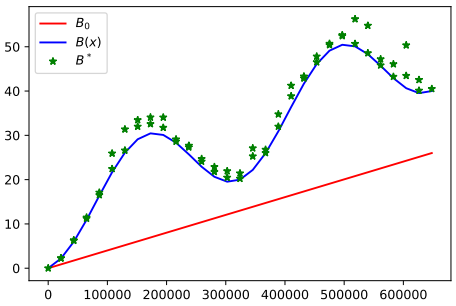


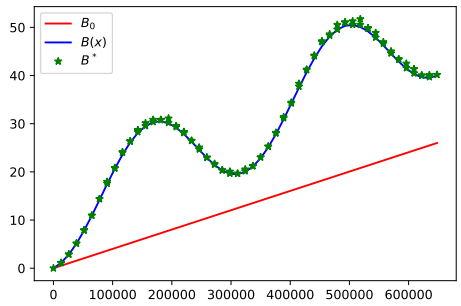
Figura 5.6: Reconstrucción de la función $\eta(x,t)$, $t = 100s, 200s, \ldots, 1000s$.

En las gráficas 5.7a y 5.7b mostramos la reconstrucción de la batimetría usando 30 y 50 elementos espaciales con un tiempo final $T_f = 1000s$ y una aproximación lineal para el método de Galerkin discontinuo. En ambos casos hemos arrancado el algoritmo de optimización con la aproximación inicial $B_0(x) = 0.75B(x)$.

Finalmente, hemos estudiado este ejemplo arrancando el algoritmo con la aproximación inicial



(a) 30 elementos espaciales.

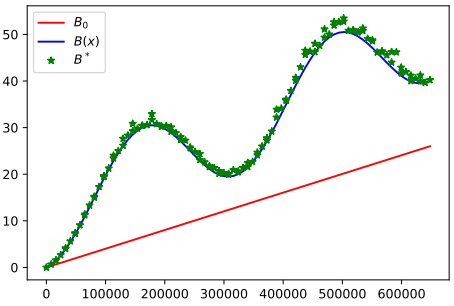


(b) 50 elementos espaciales.

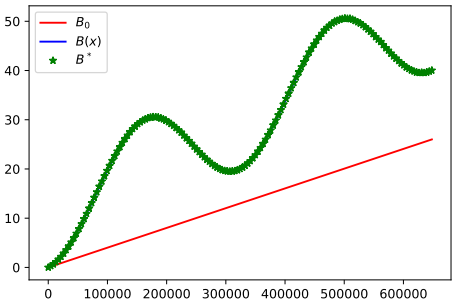
Figura 5.7: Reconstrucción de la batimetría. B_0 es la aproximación inicial, B^* es el resultado del algoritmo de optimización y $B(x)$ es la batimetría real.

$$B_0(x) = \frac{26.0}{648000.0}x,$$

y usando 40 y 75 elementos espaciales com polinomios aproximantes de grado $p = 2$. En la gráfica 5.8 mostramos los resultados obtenidos.



40 elementos espaciales.



75 elementos espaciales.

Figura 5.8: Reconstrucción de la batimetría. B_0 es la aproximación inicial, B^* es el resultado del algoritmo de optimización y $B(x)$ es la batimetría real.

En la tabla 5.2 mostramos el valor del funcional $\mathcal{J}$ en el proceso de optimización para algunas iteraciones.

Iteración k	$\mathcal{J}(B_k)$	Iteración k	$\mathcal{J}(B_k)$
3	87.266104	48	0.493704
6	37.137891	51	0.409573
9	17.332789	54	0.281501
12	11.774353	57	0.172761
15	3.633128	60	0.140771
18	1.393286	63	0.090497
21	1.360545	66	0.051329
24	1.326605	69	0.022607
27	1.265810	72	0.006300
30	1.032533	75	0.004459
33	0.945257	78	0.002612
36	0.796913	81	0.000355
39	0.604461	84	0.000041
42	0.569340	87	0.000002
45	0.542416	90	0.000000

Cuadro 5.2: Valor del funcional $\mathcal{J}$ durante el proceso de optimización.

5.5. Batimetría Discontinua

En este caso hemos considerado el espacio-temporal $[0, 1500] \times [0, 100]$ con condiciones iniciales

$$h_0(x) = 16.0 - B(x), \quad u_0(x) = 0.0,$$

donde

$$B(x) = 8\left[H(x - 562.5) - H(x - 937.5)\right], \tag{5.2}$$

con $H(x)$ la función de Heaviside, y condiciones de frontera

$$h(0, t) = 20 - 4\sin\left\{\pi\left(\frac{4t}{86400} + \frac{1}{2}\right)\right\}, \quad u(1500.0, t) = 0.$$

Este problema también es muy referido en la literatura, se puede consultar [22].

Para este ejemplo hemos usado una aproximación inicial descrita en término de funciones tipo "flan"que aproximan suavemente a la función $H(x)$, es decir, sea f la función descrita como

$$f(x; a, b, c, r) = \frac{a\exp(cr) + b\exp(rx)}{\exp(cr) + \exp(rx)},$$

donde $a < b$ corresponden a los limites entre $y = a$ y $y = b$; $r \geq 0$ cambia la pendiente y c actúa como el corrimiento en el eje x. En nuestro caso $a = 0$ y $b = 8$, de esta manera,

hemos construido las aproximaciones utilizadas como

$$B_0(x) = f(x; 0, 8, 562.5, r) - f(x; 0, 8, 937.5, r). \tag{5.3}$$

En las gráficas 5.9, mostramos el comportamiento de $B_0(x)$ con distintos valores de r comparados con la batimetría (5.2).

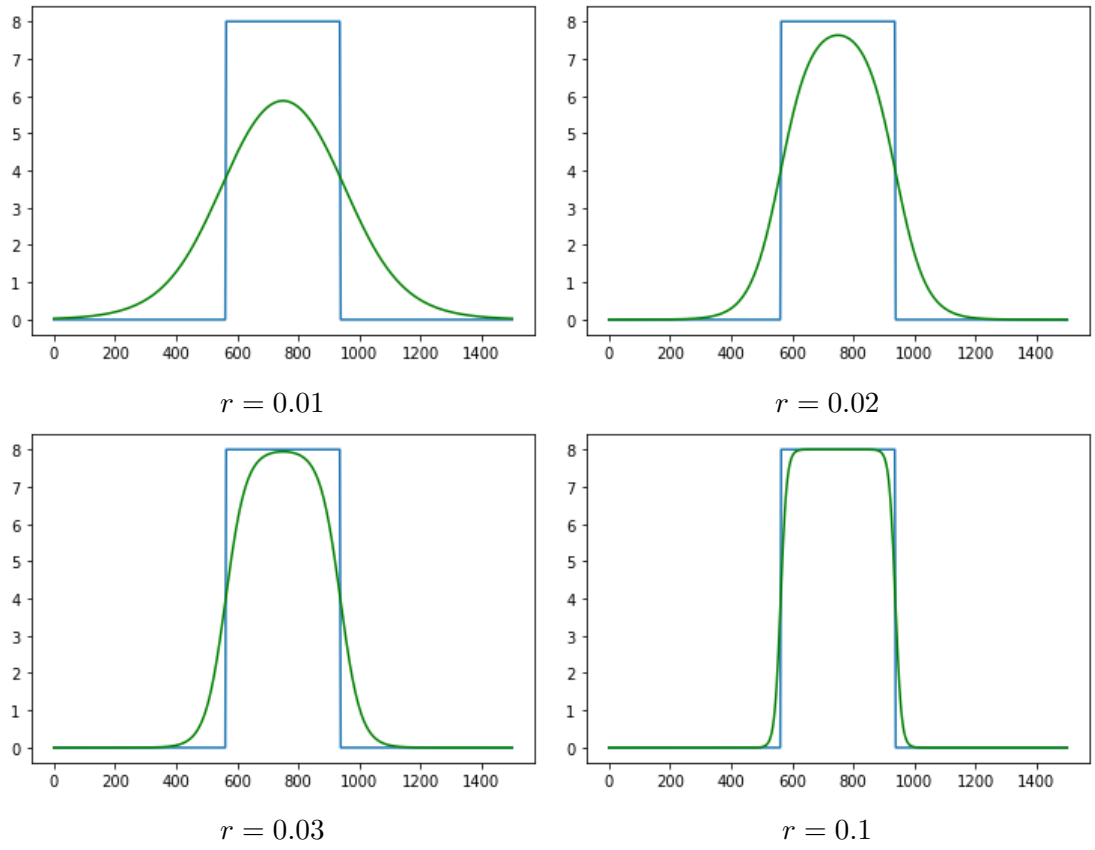
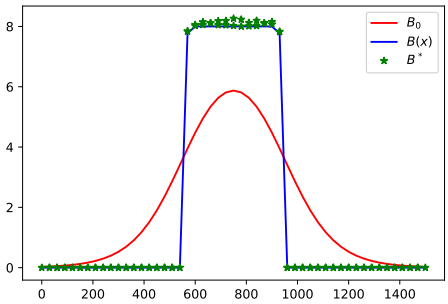
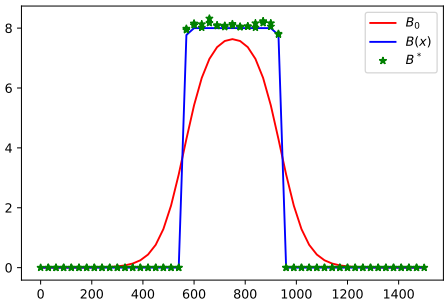


Figura 5.9: Aproximaciones a la batimetría (5.2) mediante funciones tipo "flan".

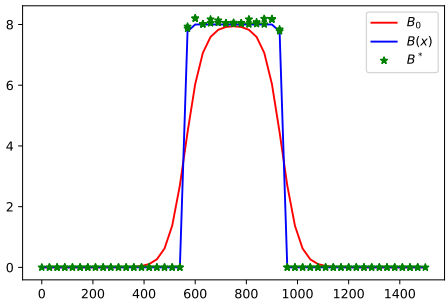
En las gráficas 5.10 podemos observar los resultados del algoritmo de optimización con 50 elementos espaciales y usando polinomios lineales en el esquema de Galerkin discontinuo.



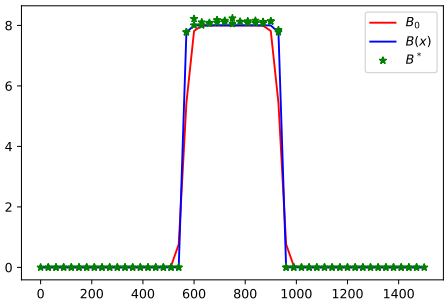
$r = 0.01$



$r = 0.02$



$r = 0.03$



$r = 0.1$

Figura 5.10: Reconstrucción de la batimetría (5.2) con distintas aproximaciones iniciales $B_0(x)$ descritas en (5.3). B^* es la aproximación obtenida mediante el algoritmo de optimización.

Capítulo 6

Sobre la Extensión al Caso Bidimensional

La metodología de la ecuación adjunta es desde luego aplicable de manera más general. En este capítulo extenderemos el método de descenso continuo. Formularemos el problema inverso y demostraremos el teorema de construcción del gradiente.

Consideremos las ecuaciones de aguas someras bidimensionales, es decir, el conjunto de ecuaciones

$$\begin{aligned} \frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} &= 0 \\ \frac{\partial hu}{\partial t} + hu\frac{\partial u}{\partial x} + hv\frac{\partial u}{\partial y} &= F_1(\Theta, h, u, v) \\ \frac{\partial hv}{\partial t} + hu\frac{\partial v}{\partial x} + hv\frac{\partial v}{\partial y} &= F_2(\Theta, h, u, v) \end{aligned} \tag{6.1}$$

donde $\Theta \in \mathbb{H}$, con $\mathbb{H}$ un espacio de Hilbert, es un conjunto de parámetros inherentes al problema y las funciones F_1 y F_2 corresponden a la situación que se este modelando. Nuestro dominio de interés es $\Omega \times [0, T]$, $\Omega \subset \mathbb{R}^2$ con frontera Lipschitz. Denotaremos como Γ la frontera de Ω, es decir, $\Gamma = \partial\Omega$.

Supongamos las condiciones iniciales

$$h(x, y, 0) = h_0(x, y), \quad u(x, y, 0) = u_0(x, y), \quad v(x, y, 0) = v_0(x, y), \quad (x, y) \in \Omega.$$

Las condiciones de frontera dependerán de la situación que estemos modelando. Ilustraremos un caso concreto: Supongamos que

$$\Gamma = \Gamma_1 \cup \Gamma_2, \quad \Gamma_1 \cap \Gamma_2 = \emptyset,$$

y fijemos condiciones de frontera sobre Γ_1, digamos

$$h(x,y,t) = h_B(t), \quad u(x,y,t) = u_B(t), \quad v(x,y,t) = v_B(t), \quad (x,y) \in \Gamma_1.$$

En [1] presentan un esquema numérico de volumen finito para resolver las ecuaciones de aguas someras bidimensionales, presentando ejemplos con condiciones iniciales y de fronteras adecuadas.

A continuación, plantearemos el problema inverso que nos interesa. Supongamos que las componentes de la velocidad (u, v) o la altura h son medidos en un conjunto de puntos $\{(x_j, y_k, t_n)\}$, $j = 1, 2, \ldots, M_1$, $k = 1, 2, \ldots, M_2$, $n = 1, 2, \ldots, N$; es decir, existe un mapeo entre el conjunto de puntos, digamos Σ, a un espacio de Hilbert $\mathcal{H}$, esto es

$$\hat{\cdot} : \Sigma \to \mathcal{H}.$$

Por ejemplo, $\mathcal{H} = \mathbb{R}^{M_1 \times M_2 \times N}$ y $\hat{\cdot}$ la función que a cada punto de Σ le asigna una medición de h, u o v y la organiza en una matriz, es decir,

$$\hat{h} = \{\hat{h}_{j,k,n}\}, \quad \hat{u} = \{\hat{u}_{j,k,n}\}, \quad \hat{v} = \{\hat{v}_{j,k,n}\}.$$

Sean $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3 : \mathcal{D} \subset \mathbb{L}^2(\Omega \times [0,T]) \to \mathcal{H}$ operadores lineales de observación. El problema de nuestro interés es estimar el conjunto de parámetros Θ, dados los datos de altura o velocidad. Para este propósito consideramos el funcional de mínimos cuadrados, $\mathcal{J} : \mathbb{H} \to \mathbb{R}$ definido de la siguiente manera:

$$\mathcal{J}(\Theta) = \frac{\alpha}{2} \left\| \mathcal{M}_1 h - \hat{h} \right\|_{\mathcal{H}}^2 + \frac{\beta}{2} \left\| \mathcal{M}_2 u - \hat{u} \right\|_{\mathcal{H}}^2 + \frac{\gamma}{2} \left\| \mathcal{M}_3 v - \hat{v} \right\|_{\mathcal{H}}^2.$$

Nuestro objetivo es minimizar $\mathcal{J}$ restringido a que h, u y v resuelven las ecuaciones de aguas someras (6.1). Los coeficientes α, β y γ son 0 o 1 dependiendo de la disponibilidad de los datos medidos. Para tal propósito, usaremos un método de descenso continuo, que ya ha sido descrito en las secciones anteriores. Recordemos que la esencia del método descrito es hallar una expresión analítica del gradiente que podamos implementar numéricamente.

Consideremos el operador Lagrangiano

$$\begin{aligned} \mathcal{L}(\Theta, h, u, v, \lambda, \mu, \sigma) &= \frac{\alpha}{2} \left\| \mathcal{M}_1 h - \hat{h} \right\|_{\mathcal{H}}^2 + \frac{\beta}{2} \left\| \mathcal{M}_2 u - \hat{u} \right\|_{\mathcal{H}}^2 + \frac{\gamma}{2} \left\| \mathcal{M}_3 v - \hat{v} \right\|_{\mathcal{H}}^2 \\ &+ \left\langle \begin{pmatrix} \lambda \\ \mu \\ \sigma \end{pmatrix}, \begin{pmatrix} \partial_t h + \partial_x (hu) + \partial_y (hv) \\ \partial_t (hu) + hu\partial_x u + hv\partial_y u - F_1(\Theta, h, u, v) \\ \partial_t (hv) + hu\partial_x v + hv\partial_y v - F_2(\Theta, h, u, v) \end{pmatrix} \right\rangle_{\mathbb{L}^2(\Omega \times [0,T])}, \end{aligned}$$

donde λ, μ y σ son los multiplicadores de Lagrange. Aquí $\langle \cdot, \cdot \rangle_{\mathbb{L}^2(\Omega \times [0,T])}$ es el producto

interior normalizado en $\mathbb{L}^2(\Omega \times [0,T])$, esto es,

$$\langle f, g \rangle_{\mathbb{L}^2(\Omega\times[0,T])} = \frac{1}{|\Omega|T} \int_0^T \int_\Omega f(x,y,t) g(x,y,t) dx dy dt.$$

Los métodos numéricos de optimización usualmente requieren encontrar la dirección de mayor decrecimiento del funcional en cada paso, el siguiente teorema nos proporciona una expresión analítica para calcular el gradiente del funcional $\mathcal{J}$.

Teorema 6.0.1 *Sean h, u y v soluciones de las ecuaciones de aguas someras con conjunto de parámetros Θ. Supongamos que los multiplicadores de Lagrange son soluciones de las ecuaciones adjuntas*

$$\begin{aligned}
\partial_t\lambda + u\partial_t\mu + v\partial_t\sigma + (\nabla\lambda - \mu\nabla u - \sigma\nabla v)\cdot(u,v) + \mu\partial_2 F_1 + \sigma\partial_2 F_2 &= \alpha\mathcal{M}_1^*(\mathcal{M}_1 h - \hat{h}),\\
h\partial_t\mu + h\partial_x\lambda - \mu h\partial_x u - \sigma h\partial_x v + \nabla\cdot\mu h(u,v) + \mu\partial_3 F_1 + \sigma\partial_3 F_2 &= \beta\mathcal{M}_2^*(\mathcal{M}_2 u - \hat{u}),\\
h\partial_t\sigma + h\partial_y\lambda - \mu h\partial_y u - \sigma h\partial_y v + \nabla\cdot\sigma h(u,v) + \mu\partial_4 F_1 + \sigma\partial_4 F_2 &= \gamma\mathcal{M}_3^*(\mathcal{M}_3 v - \hat{v}),
\end{aligned}$$

con condiciones terminales

$$\lambda(x,y,T) = \mu(x,y,T) = \sigma(x,y,T) = 0, \quad (x,t) \in \Omega,$$

y de frontera

$$\lambda(x,y,t) = \mu(x,y,t) = \sigma(x,y,t) = 0, \quad (x,y,t) \in \Gamma_2 \times [0,T).$$

Entonces la derivada de Fréchet del funcional $\mathcal{J}$ es

$$D\mathcal{J}(\Theta)\xi = -\langle \mu, DF_1(\Theta,h,u,v)\xi\rangle - \langle \sigma, DF_2(\Theta,h,u,v)\xi\rangle, \tag{6.2}$$

y de esta manera,

$$\nabla\mathcal{J}(\Theta) = -(DF_1(\Theta,h,u,v))^*\mu - (DF_2(\Theta,h,u,v))^*\sigma. \tag{6.3}$$

Demostración: La demostración la dividiremos en tres pasos:

1. Calculo de la derivada del operador $\mathcal{L}$.
2. Observemos que cuando $(h(\Theta), u(\Theta), v(\Theta))$ son soluciones de las ecuaciones de aguas someras para el conjunto de parámetros Θ, entonces el operador $\mathcal{L}$ coincide con $\mathcal{J}$ y aplicaremos la regla de la cadena para obtener la derivada de Fréchet de $\mathcal{J}(\Theta)$.
3. Usaremos propiedades de la derivada de Fréchet para establecer condiciones iniciales para $Dh(\Theta)\xi$, $Du(\Theta)\xi$, y $Dv(\Theta)\xi$; y aplicaremos integración por partes para construir el operador adjunto de $\mathcal{T} = \partial_i$, para $i = x, y, t$ con el propósito de reescribir $D\mathcal{J}(\Theta)\xi$.

Nota: Usaremos la notación D_i para las derivadas de Fréchet y ∂_i para las derivadas parciales de funciones en el sentido usual.

Paso 1: La derivada del operador $\mathcal{L}$ está dada por

$$\begin{aligned} D\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)(\xi,\eta,\zeta,\varsigma) &= D_1\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\xi + D_2\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\eta \\ &+ D_3\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\zeta + D_4\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\varsigma. \end{aligned}$$

A continuación calcularemos cada uno de los términos de la expresión anterior.

$$\begin{aligned} D_1\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\xi &= D\left\langle\mu,-F_1(\cdot,h,u,v)\right\rangle(\Theta)\xi + D\left\langle\sigma,-F_2(\cdot,h,u,v,w)\right\rangle(\Theta)\xi, \\ D_2\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\eta &= \left\langle\eta,\alpha\mathcal{M}_1^*(\mathcal{M}_1h-\hat{h})\right\rangle + \langle\lambda,\eta_t+\nabla\cdot\eta(u,v)\rangle \\ &+ \langle\mu,(\eta u)_t+\eta\nabla u\cdot(u,v)-\partial_2F_1(\Theta,h,u,v)\eta\rangle \\ &+ \langle\sigma,(v\eta)_t+\eta\nabla v\cdot(u,v)-\partial_2F_2(\Theta,h,u,v)\eta\rangle, \\ D_3\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\zeta &= \langle\zeta,\beta\mathcal{M}_2^*(\mathcal{M}_2u-\hat{u})\rangle + \langle\lambda,\partial_x(h\zeta)\rangle \\ &+ \langle\mu,(h\zeta)_t+\nabla\zeta\cdot h(u,v)+h\zeta\partial_xu-\partial_3F_1(\Theta,h,u,v)\zeta\rangle \\ &+ \langle\sigma,h\zeta\partial_xv-\partial_3F_2(\Theta,h,u,v)\zeta\rangle, \\ D_4\mathcal{L}(\Theta,h,u,v,\lambda,\mu,\sigma)\varsigma &= \langle\varsigma,\gamma\mathcal{M}_3^*(\mathcal{M}_3v-\hat{v})\rangle + \langle\lambda,\partial_y(h\varsigma)\rangle \\ &+ \langle\mu,h\varsigma\partial_yu-\partial_4F_1(\Theta,h,u,v)\varsigma\rangle \\ &+ \langle\sigma,(h\varsigma)_t+\nabla\varsigma\cdot h(u,v)+h\varsigma\partial_yv-\partial_4F_2(\Theta,h,u,v)\varsigma\rangle. \end{aligned}$$

Paso 2: Sea $W(\Theta)=(\Theta,h(\Theta),u(\Theta),v(\Theta))$, donde $(h(\Theta),u(\Theta),v(\Theta)$ resuelven las ecuaciones de aguas someras para un conjunto de parámetros dados. Entonces

$$\mathcal{J}(\Theta)=\mathcal{L}(W(\Theta)).$$

Al usar la regla de la cadena, obtenemos

$$\begin{aligned} D\mathcal{J}(\Theta)\xi &= D\mathcal{L}(W(\Theta))DW(\Theta)\xi \\ &= D\mathcal{L}(W(\Theta))(\xi,Dh(\Theta)\xi,Du(\Theta)\xi,Dv(\Theta)\xi) \\ &\equiv D\mathcal{L}(W(\Theta))(\xi,H,U,V). \end{aligned}$$

De esta manera,

$$\begin{aligned} D\mathcal{J}(\Theta)\xi &= D\left\langle \mu, -F_1(\cdot,h,u,v)\right\rangle(\Theta)\xi + D\left\langle \sigma, -F_2(\cdot,h,u,v)\right\rangle(\Theta)\xi \\ &+ \left\langle H, \alpha\mathcal{M}_1^*(\mathcal{M}_1 h - \hat{h})\right\rangle + \langle \lambda, H_t + \nabla\cdot H(u,v)\rangle \\ &+ \langle \mu, (Hu)_t + H\nabla u\cdot(u,v) - \partial_2 F_1(\Theta,h,u,v)H\rangle \\ &+ \langle \sigma, (Hv)_t + H\nabla v\cdot(u,v) - \partial_2 F_2(\Theta,h,u,v)H\rangle \\ &+ \langle U, \beta\mathcal{M}_2^*(\mathcal{M}_2 u - \hat{u})\rangle + \langle \lambda, \partial_x(Uh)\rangle \\ &+ \langle \mu, (Uh)_t + \nabla U\cdot h(u,v) + hU\partial_x u - \partial_3 F_1(\Theta,h,u,v)U\rangle \\ &+ \langle \sigma, hU\partial_x v - \partial_3 F_2(\Theta,h,u,v)U\rangle + \langle V, \gamma\mathcal{M}_3^*(\mathcal{M}_3 v - \hat{v})\rangle \\ &+ \langle \lambda, \partial_y(Vh)\rangle + \langle \mu, hV\partial_y u - \partial_4 F_1(\Theta,h,u,v)V\rangle \\ &+ \langle \sigma, (hV)_t + \nabla V\cdot h(u,v) + hV\partial_y v - \partial_4 F_2(\Theta,h,u,v)V\rangle\,. \end{aligned}$$

Paso 3: Ya que la diferenciabilidad Fréchet implica Gâteaux diferenciabilidad (y amabas derivadas coinciden), entonces

$$Df(\Theta)\xi = \lim_{\varepsilon\to 0}\frac{f(\Theta+\varepsilon\xi) - f(\Theta)}{\varepsilon},$$

para $f = h, u, v$.

Estas expresiones implican que las funciones H, U y V se anulan en los puntos donde tenemos condiciones iniciales y de frontera, en otras palabras

$$\begin{aligned} &H(x,y,0) = U(x,y,0) = V(x,y,0) = 0, \quad (x,y)\in\Omega, \\ &H(x,y,t) = U(x,y,t) = V(x,y,t) = 0, \quad (x,y,t)\in\Gamma_1\times(0,T]. \end{aligned} \tag{6.4}$$

Usaremos las condiciones iniciales y de fronteras (6.4) para reescribir la expresión de $D\mathcal{J}(\Theta)\xi$ mediante integración por partes (construcción de operadores adjuntos).

Sea $\nu = (\nu^{(1)}, \nu^{(2)}, \nu^{(3)})$ el vector normal unitario exterior a la frontera del conjunto $\Omega\times[0,T]$, y consideremos el operador diferencial $\mathcal{T} = \partial_i$ para $i = x, y, t$. Sean f y g funciones tales que

$$\begin{aligned} f(x,y,t) &= 0, \quad (x,y,t)\in\Omega\times\{T\} \quad f(x,y,t) = 0 \quad (x,y,t)\in\Gamma_2\times[0,T] \\ g(x,y,t) &= 0, \quad (x,y,t)\in\Omega\times\{0\} \quad g(x,y,t) = 0 \quad (x,y,t)\in\Gamma_1\times[0,T]. \end{aligned} \tag{6.5}$$

Entonces

$$\begin{aligned}
|\Omega|T\left\langle f,\mathcal{T}g\right\rangle &= \int\limits_{\Omega\times[0,T]} f\mathcal{T}g d\mathbf{x} = -\int\limits_{\Omega\times[0,T]} g\mathcal{T}f d\mathbf{x} + \int\limits_{\partial(\Omega\times[0,T])} fg\nu^i dS \\
&= |\Omega|T\left\langle -\mathcal{T}f,g\right\rangle + \int\limits_{\partial\Omega\times[0,T]} fg\nu^i dS + \int\limits_{\Omega\times\{0,T\}} fg\nu^i dS \\
&= |\Omega|T\left\langle -\mathcal{T}f,g\right\rangle + \int\limits_{(\Gamma_1\cup\Gamma_2)\times[0,T]} fg\nu^i dS + \int\limits_{\Omega\times\{0,T\}} fg\nu^i dS \\
&= |\Omega|T\left\langle -\mathcal{T}f,g\right\rangle + \int\limits_{\Gamma_1\times[0,T]} fg\nu^i dS + \int\limits_{\Gamma_2\times[0,T]} fg\nu^i dS + \int\limits_{\Omega\times\{0,T\}} fg\nu^i dS \\
&= |\Omega|T\left\langle -\mathcal{T}f,g\right\rangle .
\end{aligned}$$

Por lo tanto,

$$\left\langle f,\mathcal{T}g\right\rangle = \left\langle -\mathcal{T}f,g\right\rangle ,$$

en otras palabras $\mathcal{T}^* = -\mathcal{T}$.

Observemos que los multiplicadores de Lagrange (λ,μ,σ) y las funciones H, U V cumplen las condiciones de f y g en (6.5), respectivamente. De esta manera podemos escribir

$$\begin{aligned}
\left\langle \lambda,\partial_t H+\nabla\cdot H(u,v)\right\rangle + \left\langle \mu,\partial_t(Hu)\right\rangle + \left\langle \sigma,\partial_t(Hv)\right\rangle &= \left\langle H,-\partial_t\lambda-\nabla\lambda\cdot(u,v)-u\partial_t\mu-v\partial_t\sigma\right\rangle , \\
\left\langle \lambda,\partial_x(Uh)\right\rangle + \left\langle \mu,\partial_t(Uh)+\nabla U\cdot h(u,v)\right\rangle &= \left\langle U,-h\partial_x\lambda-h\partial_t\mu-\nabla\cdot\mu h(u,v)\right\rangle , \\
\left\langle \lambda,\partial_y(Vh)\right\rangle + \left\langle \sigma,\partial_t(hV)+\nabla V\cdot h(u,v)\right\rangle &= \left\langle V,-h\partial_y\lambda-h\partial_t\sigma-\nabla\cdot\sigma h(u,v)\right\rangle .
\end{aligned}$$

Así,

$$\begin{aligned}
D\mathcal{J}(\Theta)\xi &= D\left\langle \mu,-F_1(\cdot,h,u,v)\right\rangle(\Theta)\xi + D\left\langle \sigma,-F_2(\cdot,h,u,v)\right\rangle(\Theta)\xi \\
&+ \left\langle H,\alpha\mathcal{M}_1^*(\mathcal{M}_1h-\hat{h})\right\rangle + \left\langle H,-\partial_t\lambda-\nabla\lambda\cdot(u,v)\right\rangle \\
&+ \left\langle H,-u\partial_t\mu+\mu\nabla u\cdot(u,v)-[\partial_2F_1(\Theta,h,u,v)]\mu\right\rangle \\
&+ \left\langle H,-v\partial_t\sigma+\sigma\nabla v\cdot(u,v)-[\partial_2F_2(\Theta,h,u,v)]\sigma\right\rangle \\
&+ \left\langle U,\beta\mathcal{M}_2^*(\mathcal{M}_2u-\hat{u})\right\rangle + \left\langle U,-h\partial_x\lambda\right\rangle \\
&+ \left\langle U,-h\partial_t\mu-\nabla\cdot\mu h(u,v)+\mu h\partial_x u-[\partial_3F_1(\Theta,h,u,v)]\mu\right\rangle \\
&+ \left\langle U,\sigma h\partial_x v-[\partial_3F_2(\Theta,h,u,v)]\sigma\right\rangle + \left\langle V,\gamma\mathcal{M}_3^*(\mathcal{M}_3v-\hat{v})\right\rangle \\
&+ \left\langle V,-h\partial_y\lambda\right\rangle + \left\langle V,\mu h\partial_y u-[\partial_4F_1(\Theta,h,u,v)]\mu\right\rangle \\
&+ \left\langle V,-h\partial_t\sigma-\nabla\cdot\sigma h(u,v)+\sigma h\partial_y v-[\partial_4F_2(\Theta,h,u,v)]\sigma\right\rangle .
\end{aligned}$$

Por lo tanto,

$$D\mathcal{J}(\Theta)\xi = -\langle \mu, DF_1(\Theta, h, u, v)\xi\rangle - \langle \sigma, DF_2(\Theta, h, u, v)\xi\rangle .$$

Sean $\mathcal{K}_i \equiv (DF_i(\Theta, h, u, v))^*$, $i = 1, 2$; entonces

$$D\mathcal{J}(\Theta)\xi = -\langle \mathcal{K}_1\mu, \xi\rangle - \langle \mathcal{K}_2\sigma, \xi\rangle = \langle -\mathcal{K}_1\mu - \mathcal{K}_2\sigma, \xi\rangle .$$

De la definición 1.2.18 (Gradiente de un funcional, o equivalentemente, del Teorema de Representación de Riesz 1.2.17), podemos escribir

$$\nabla\mathcal{J}(\Theta) = -\mathcal{K}_1\mu - \mathcal{K}_2\sigma,$$

como se quería. □

Al aplicar el método de Galerkin Discontinuo a las ecuaciones adjuntas, surge el problema de evaluar integrales con la forma

$$\int_{\Omega_i} (\mathcal{M}^* f)(x, y, t)\varphi(x, y)dxdy,$$

para $f \in \mathcal{H}$, $\varphi(x, y)$ una función prueba y $\Omega_i \subset \Omega$. Para tal propósito, usaremos el Teorema de Diferenciación de Lebesgue (Teorema 3.3.1):

Sea $\Delta t > 0$ y consideremos el conjunto $E_t = [t - \Delta t/2, t + \Delta t/2]$, entonces

$$\begin{aligned}
\int_{\Omega_i} (\mathcal{M}^* f)(x, y, t)\varphi(x, y)dxdy &= \int_{\Omega} (\mathcal{M}^* f)(x, y, t)\varphi(x, y)\chi_{\Omega_i}(x, y)dxdy \\
&\approx \frac{1}{\Delta t}\int_{t-\Delta t/2}^{t+\Delta t/2}\int_{\Omega} (\mathcal{M}^* f)(x, y, t)\varphi(x, y)\chi_{\Omega_i}(x, y)dxdydt \\
&= \frac{1}{\Delta t}\int_0^T\int_{\Omega} (\mathcal{M}^* f)(x, y, t)\varphi(x, y)\chi_{\Omega_i}(x, y)\chi_{E_t}(t)dxdydt \\
&= \frac{|\Omega|T}{\Delta t}\left\langle (\mathcal{M}^* f)(x, y, t), \varphi(x, y)\chi_{\Omega_i}(x, y)\chi_{E_t}(t)\right\rangle_{\mathbb{L}^2(\Omega\times[0,T])} \\
&= \frac{|\Omega|T}{\Delta t}\left\langle f, \mathcal{M}(\varphi(x, y)\chi_{\Omega_i}(x, y)\chi_{E_t}(t))\right\rangle_{\mathcal{H}} .
\end{aligned}$$

La expresión del gradiente depende del modelo subyacente, para ejemplificar el proceso, estudiaremos dos casos concretos.

Ejemplo 6.0.1 *Si consideramos la batimetría en el término fuente, entonces podemos establecer para* $\Theta = B(x,y)$

$$\begin{bmatrix} F_1(B,h,u,v) \\ F_2(B,h,u,v) \end{bmatrix} = -gh\nabla B,$$

donde g *es la constante gravitacional y* ∇B *es la pendiente de la batimetría.*

Consideremos $\xi \in \mathcal{C}_c^\infty(\Omega)$. *Ya que el operador derivada es lineal, entonces*

$$DF_1(B,h,u,v)\xi = -gh\partial_x\xi, \quad DF_2(B,h,u,v)\xi = -gh\partial_y\xi.$$

Así,

$$D\mathcal{J}(\Theta)\xi = \langle \partial_x\xi, gh\mu\rangle + \langle \partial_y\xi, gh\sigma\rangle .$$

Nota: Recordemos que $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{\mathbb{L}^2(\Omega\times[0,T])}$.

Al usar integración por partes,

$$D\mathcal{J}(\Theta)\xi = \left\langle \xi, -g\int_0^T \partial_x(h\mu)dt \right\rangle_{\mathbb{L}^2(\Omega)} + \left\langle \xi, -g\int_0^T \partial_y(h\sigma)dt \right\rangle_{\mathbb{L}^2(\Omega)}.$$

Por lo tanto,

$$\nabla\mathcal{J}(B) = -g\int_0^T \nabla\cdot h(\mu,\sigma)dt.$$

Si suponemos que tenemos mediciones de las componentes de velocidad (u,v), *entonces las ecuaciones adjuntas toman la forma*

$$\begin{aligned} \partial_t\lambda + u\partial_t\mu + v\partial_t\sigma + (\nabla\lambda - \mu\nabla u - \sigma\nabla v)\cdot(u,v) - g\nabla B\cdot(\mu,\sigma) &= 0, \\ h\partial_t\mu + h\partial_x\lambda - \mu h\partial_x u - \sigma h\partial_x v + \nabla\cdot\mu h(u,v) &= \mathcal{M}_2^*(\mathcal{M}_2 u - \hat{u}), \\ h\partial_t\sigma + h\partial_y\lambda - \mu h\partial_y u - \sigma h\partial_y v + \nabla\cdot\sigma h(u,v) &= \mathcal{M}_3^*(\mathcal{M}_3 v - \hat{v}). \end{aligned}$$

Ejemplo 6.0.2 *Si consideremos la batimetría* B *y el coeficiente de fricción de Manning* κ *en el término fuente, entonces para* $\Theta = (B,\kappa)$

$$\begin{bmatrix} F_1(\Theta,h,u,v) \\ F_2(\Theta,h,u,v) \end{bmatrix} = gh(S_0 - S_h), \quad S_0 = -\nabla B, \quad S_h = \frac{\kappa\sqrt{u^2+v^2}}{h^{4/3}}\begin{bmatrix} u \\ v \end{bmatrix}.$$

Sea $\Phi = (\xi,\eta)$, *tal que* $\xi \in \mathcal{C}_c^\infty(\Omega)$ *y* $\eta \in \mathbb{R}$. *Entonces,*

$$\begin{aligned} DF_1(B,\kappa,h,u,v)\Phi &= D_1F_1(B,\kappa,h,u,v)\xi + D_2F_1(B,\kappa,h,u,v)\eta, \\ DF_2(B,\kappa,h,u,v)\Phi &= D_1F_2(B,\kappa,h,u,v)\xi + D_2F_2(B,\kappa,h,u,v)\eta. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} DF_1(B,\kappa,h,u,v)\Phi &= -gh\partial_x\xi - \frac{gu}{h^{1/3}}\sqrt{u^2+v^2}\eta, \\ DF_2(B,\kappa,h,u,v)\Phi &= -gh\partial_y\xi - \frac{gv}{h^{1/3}}\sqrt{u^2+v^2}\eta. \end{aligned}$$

Haciendo manipulaciones como en el ejemplo anterior, podemos escribir

$$D\mathcal{J}(B,\kappa)\Phi = \left\langle \xi, -g\int_0^T \nabla\cdot h(\mu,\sigma)dt \right\rangle_{\mathbb{L}^2(\Omega)} + \eta\left\langle u\mu + v\sigma, \frac{g\sqrt{u^2+v^2}}{h^{1/3}} \right\rangle.$$

Entonces,

$$\nabla\mathcal{J}(B,\kappa) = \left(-g\int_0^T \nabla\cdot h(\mu,\sigma)dt, \left\langle u\mu + v\sigma, \frac{g\sqrt{u^2+v^2}}{h^{1/3}} \right\rangle\right) \in \mathbb{L}^2(\Omega)\times\mathbb{R}.$$

Capítulo 7

Conclusiones y Trabajo Futuro

En este trabajo hemos considerado el problema inverso de estimación de parámetros en ecuaciones de aguas someras. En particular, la estimación de la batimetría cuando se tienen mediciones de velocidad. De nuestro conocimiento este problema no se ha tratado en la literatura. La motivación del estudio es la reciente publicación [5], donde se introduce técnicas de medición de la velocidad en flujos superficiales.

Se ha desarrollado la metodología de la ecuación adjunta para la solución. Con técnicas de Análisis Funcional No lineal y Teoría de Ecuaciones Diferenciales Parciales, se ha construido el gradiente del funcional a minimizar. Este gradiente es un elemento esencial en la implementación del método del descenso del gradiente continuo.

Desde el punto de vista numérico, se ha desarrollado e implantado la solución en base al método de Galerkin Discontinuo. Este método ha cobrado recientemente gran importancia en la solución numérica de sistemas hiperbólicos no lineales. En el problema que nos concierne, una ventaja, es la aproximación natural al operador adjunto del operador de observación en la formulación débil del sistema hiperbólico adjunto.

La metodología se ha aplicado a dos problemas de gran interés, pero no abordados de manera completa en la literatura. Esto son, recuperación de batimetría en condiciones de mareas oceánicas, y recuperación de batimetrías discontinuas. Consideramos dos problemas propuestos en la literatura, donde solo se aborda el problema directo. Nuestros resultados son muy satisfactorios. Por completes, incluimos también recuperación de casos clásicos de referencia.

El método de estimación de batimetría es un método iterativo que requiere una estimación inicial. El desempeño depende enormemente en esta elección por ser un método local. Nuestra ampliación a la recuperación de la batimetría en situación de mareas oceánicas, muestra la robustez de nuestro enfoque. La estimación de la batimetría es precisa,

a pesar de iniciar con una estimación inicial *lejana*.

Como continuación de este trabajo de tesis, existen diversas líneas de investigación que quedan abiertas. Como mostramos en el capítulo 6, la metodología teórica es aplicable al caso bidimensional. El complemento numérico-computacional queda por desarrollarse.

Un problema abierto de gran interés teórico, es uno de identificabilidad, esto es, determinar la inyectividad del mapeo de parámetros a datos. De nuestro conocimiento, los avances son a lo mas incipientes en el caso de las ecuaciones de aguas someras.

También para la robustez del método, es preciso explorar resultados de sensibilidad. El conocimiento del gradiente del funcional es un primer paso en tal dirección.

Bibliografía

[1] Anastasiou, K. - Chan, C. T. (1997). Solution of the 2D Shallow Water Equations Using the Finite Volume Method on Unstructured Triangular Meshes. *International Journal for Numerical Methods in Fluids*, 24(11):1225–1245.

[2] Aragón, F. - Goberna, M. - López, M. - Rodríguez, M. (2019). *Nonlinear Optimization*. Springer Verlag, first edition.

[3] Barth, Timothy J. (2003). *High-Order Methods for Computational Physics*. Springer, second edition.

[4] Bermúdez A. Vázquez, M. E. (1994). Upwind Methods for Hyperbolic Conservation Laws with Source Terms. *Elsevier*, 23(8):1049–1071.

[5] Bolognesi, M. - Farina, G. - Alvisi, S. - Franchini, M. - Pellegrinelli, A. - Russo, P. (2017). Measurement of Surghtweight Velocity in Open Channels Using a Lightweight Remotely Piloted Aircraft System. *Geomatics, Natural Hazards and Risk*, 8(1):73–86.

[6] Boyd, S. (2011). *Convex Optimization*. Cambridge University Press.

[7] Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer-Verlag.

[8] Cheney, W. (2001). *Analysis for Applied Mathematics*. Springer-Verla.

[9] Cockburn, B. Shu, C. W. (1989). TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation laws II: General Framework. *Mathematics of Computation*, 52(89):411–435.

[10] Dieudonné, J. (1960). *Foundations of Modern Analysis*. Academic Press.

[11] Evans, L. C. (2010). *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, second edition.

[12] Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications.* John Wiley and Sons.

[13] Gessese, A. - Sellier, M. - Van Houten, E. - Smart, G. (2011). Reconstruction of River Bed Topography From Free Surface Data Using a Direct Numerical Approach in One-Dimensional Shallow Water Flow. *Inverse Problems*, 27(2):025001.

[14] Gessese, A. - Smart, G. - Heining, C. - Sellier, M. (2013). One-Dimensional Bathymetry Based on Velocity Measurements. *Inverse Problems in Science and Engineering*, 21(4):704–720.

[15] Gottlieb, S. - Ketcheson, D. - Shu, C. (2011). *Strong Stability Preserving Runge-Kutta and Multistep Time Discretizations.* World Scientific.

[16] Harten, A. (1983). High Resolution Schemes for Hyperbolic Conservation Laws. *Journal of Computational Physics*, 49(3):357–393.

[17] Heath, M. T. (1996). *Scientific Caomputing. An Introductory Survey.* McGraw-Hill Companies.

[18] Hesthaven, J. - Warburton, T. (2010). *Nodal Discontinuous Galerkin Methods.* Springer Verlag, New York, first edition.

[19] Hughes, T. J. R. - Liu, W. K. - Brooks, A. (1979). Finite Element Analysis of Incompressible Viscous Flows by the Penalty Function Formulation. *Journal of Computational Physics*, 30(1):1–60.

[20] Kevlahan, N. - Khan, R. - Protas, B. (2019). On the Convergence of Data Assimilation for the One-Dimensional Shallow Water Equations with Sparse Observations. *Advances in Computational Mathematics*, 45(5):3195–3216.

[21] Krivodonova, L. (2007). Limiters for High-Order Discontinuous Galerkin Methods. *Journal of Computational Physics*, 226(1):879–896.

[22] Lee, H. (2020). Implicit Discontinuous Galerkin Scheme for Discontinuous Bathymetry in Shallow Water Equations. *KSCE Journal of Civil Engineerings*, 24(9):2694–2705.

[23] Lee, J. - Ghorbanidehno, H. - Farthing, M. Hesser, T - Darve, E. - Kitanidis, P. (2018). Riverine Bathymetry Imaging With Indirect Observations. *Water Resources Research*, 54(5):3704–3727.

[24] LeVeque, R. J. (1990). *Numerical Methods for Conservation Laws.* Springer-Verlag.

[25] LeVeque, R. L. (2011). *Finite Volume Methods for Hyperbolic Problems.* Cambridge Univ. Press, second edition.

[26] Lin, G. F. - Lai, J. S. - Guo, W. D. (2003). Finite-Volume Component-Wise TVD Schemes for 2D Shallow Water Equations. *Journal in Advances in Water Resources*, 26(8):861–873.

[27] Minoux, M. (1986). *Mathematical Programming. Theory and Algorithms.* John Wiley and Sons.

[28] Montecinos, G. (2019). A Numerical Procedure and Coupled System Formulation for the Adjoint Approach in Hyperbolic PDE-Constrained Optimization Problems. *IMA Journal of Applied Mathematics*, 84(3):483–516.

[29] Nguyen, V. (2016). State and Parameter Estimation in 1-D Hyperbolic PDEs Based on an Adjoint Method. *Automatica*, 67:185–191.

[30] Nguyen, V. - George, D. Besancon, G. (2016). Adjoint-based State and Distributed Parameter Estimation in a Switched Hyperbolic Overland Flow Model. *IFAC-PapersOnLine*, 49(18):205–210.

[31] Nguyen, V. - George, D. Besancon, G. - Zin, I. (2016). Parameter Estimation of a Real Hydrological System Using an Adjoint Method. *IFAC-PapersOnLine*, 49(13):300–305.

[32] Peressini, A. - Sullivan, F. - Uhl, J. (1996). *The Mathematics of Nonlinear Programming.* Springer Verlag, New York, second edition.

[33] Quarteroni, A. - Saleri, F. - Sacco, R. (2000). *Numerical Mathematics.* Springer Verlag, New York, first edition.

[34] Reed, W. H. Hill, T. (1973). Triangular Mesh Method for the Neutron Transport Equation. *Los Alamos Report*, LA-UR(73-479).

[35] Saint-Venant, J. (1871). Théorie du mouvement non-permanent des eaux, avec application aux crues des rivieres. *Comptes rendus de l'Académie des Sciences de Paris*, 73(4):521–528.

[36] Salsa, S. (2007). *Partial Differential Equations in Action.* Springer, second edition.

[37] Sellier, M. (2015). Inverse Problems in Free Surface Flows: A Review. *Acta Mechanica*, 227(3):913–935.

[38] Smoller, J. (1994). *Shock Waves and Reaction-Diffusion Equations.* Springer-Verlag.

[39] Tirupathi, S. - Tchrakia, T. - Zhuk, S. - McKenna, s. (2016). Shock Capturing Data Assimilation Algorithm for 1D Shallow Water Equations. *Advances in Water Resources*, 88:198–210.

[40] Van Leer, B. (1979). Towards the Ultimate Conservative Difference Scheme. V. A Second-Order Sequel to Godunov's Method. *Journal of Computational Physics*, 32(1):101–136.

[41] Wang, J. S. Ni, H. G. He, Y. S. (2000). Finite Difference TVD Scheme for Computation of Dam-Break Problems. *Journal of Hydraulic Engineering*, 126(4):253–262.

[42] Wloka, J. (1987). *Partial Differential Equations*. Cambridge University Press.

[43] Xing, Y. (2016). Numerical Methods for the Nonlinear Shallow Water Equations. *Handbook of Numerical Analysis*, 18(1):361–384.

[44] Yang, X.-S. (2019). *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press.

[45] Zienkiewicz, C. Ortiz, P. (1995). A Split-Characteristic Based Finite Element Model for the Shallow Equations. *International Journal for Numerical Methods in Fluids*, 20(8-9):1061–1080.