UNIVERSIDAD DE EL SALVADOR

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA DE MATEMÁTICA

LICENCIATURA EN MATEMÁTICA

Universidad de El Salvador

Hacia la libertad por la cultura

TRABAJO DE GRADUACIÓN TITULADO:

MÉTODOS NUMÉRICOS PARA ECUACIONES DIFERENCIALES PARCIALES

PREVIO A OBTENER EL TÍTULO DE:

LICENCIADO EN MATEMÁTICA

PRESENTADO POR:

Br. Willian Armando Miranda Tobar. *Carné:* MT09001

DOCENTE ASESOR:

MSc. Carlos Ernesto Gámez Rodríguez.

Ciudad Universitaria, Noviembre de 2014.

AUTORIDADES UNIVERSITARIAS PERIODO 2011-2015

UNIVERSIDAD DE EL SALVADOR

RECTOR: ING. MARIO ROBERTO NIETO LOVO

VICERRECTORA ACADÉMICA: MAESTRA ANA MARÍA GLOWER DE ALVARADO

VICERRECTOR ADMINISTRATIVO: MAESTRO ÓSCAR NOÉ NAVARRETE

SECRETARIA GENERAL: DRA. ANA LETICIA DE AMAYA

DEFENSORA DE LOS DERECHOS UNIVERSITARIOS: LICDA. CLAUDIA MARÍA MELGAR DE ZAMBRANA

FISCAL GENERAL:LIC. FRANCISCO CRUZ LETONA

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DECANO: M.SC. MARTÍN ENRIQUE GUERRA CÁCERES

VICEDECANO: LIC. RAMÓN ARÍSTIDES PAZ SÁNCHEZ

SECRETARIO: LIC. CARLOS ANTONIO QUINTANILLA APARICIO

ESCUELA DE MATEMATICA

DIRECTOR: DR. JOSÉ NERYS FUNES TORRES

SECRETARIA: LICDA. ALBA IDALIA CÓRDOVA CUELLAR

Agradecimientos

Quiero agradecer en primer lugar a Dios, por estar presente en todos los momentos del transcurso de mi vida, en especial durante mi carrera, por permitir que terminara con éxito.

También agradezco de todo corazón a mis padres: Maximiliano Miranda y Leticia Tobar, quienes me brindaron todo el apoyo posible y a quienes debo lo que soy. Por ser ejemplo de honorabilidad, perseverancia y esfuerzo. Y quienes en todo momento me acompañaban ya sea como padres, consejeros o amigos. Ellos saben que no me alcanzaría esta página para agradecerles y para decirles lo mucho que agradezco a mi Dios por ponerlos ahí, siempre junto a mí.

A mis hermanos: Jessenia, Tony, Nixon y Balmoré, por ser unos excelentes hermanos, por estar siempre conmigo, por todo el apoyo recibido de ellos y darme grandes lecciones.

A mi asesor, Carlos Gámez, a quien aprecio mucho. Le agradezco infinitamente su paciencia, su buena disposición y apoyo durante el desarrollo del Trabajo de Graduación. Gran parte de lo que he logrado en esta etapa ha sido, gracias al respaldo de él. Un verdadero gusto trabajar bajo su dirección.

A los profesores de la Escuela de Matemática de la UES, que me brindaron de su tiempo y conocimiento. En particular al Msc. Porfirio Rodríguez por revisar el proyecto inicial y al Dr. Simón Peña y al Lic. Edwin Aguilar, quienes fueron parte del tribunal calificador del Trabajo de Graduación. Al Msc. Pedro Ramos, por permitirme ser parte de sus iniciativas.

Agradezco mucho el apoyo que he recibido de parte de mis tíos durante este tiempo, infinitamente, les agradezco a toda mi familia en general por esta junto a mi todo este tiempo esperando de su apoyo siempre para seguir adelante y en el camino correcto.

A mis compañeros y amigos que me acompañaron en diferentes momentos de la carrera. Y a todas las personas que he conocido durante estos años, que de una manera u otra han colaborado para que llegar aquí sea posible, a todos gracias.

Índice

1. Resumen

Este trabajo consiste en el estudio de un tema que tiene fundamental importancia en el estudio numérico de las ecuaciones diferenciales parciales, resolviendo diferentes problemas por medio del método de diferencias finitas y el método de elementos finitos provehiendo para esto códigos implementables en Octave/MATLAB.

El presente trabajo pretende ser una introducción al tema de métodos numéricos para ecuaciones diferenciales parciales. Estará dividido en dos partes: La primera dedicada al método de diferencias finitas y la otra parte al método de elementos finitos. Esta dividido en 7 capítulos.

En el Capítulo 1 se dá la teoría básica necesaria para iniciar nuestro estudio, en el Capítulo 2 se explicará el método de diferencias finitas para la solución de un problema modelo dando diferentes características de éste y en los Capítulos 3, 4 y 5 se aplicará dicho método para resolver un problema modelo de la ecuación de calor, onda y Laplace, respectivamente, escribiendo un código implementable en Octave/MATLAB.

En el Capítulo 6 se explica el método de elementos finitos en una dimensión y en el Capítulo 7 se implementa el método de elementos finitos para lo cual se analiza parte paper de J. Alberty, C. Carstensen y S.A. Funken que proporciona la implementación del método.

2. Introducción

Con las ecuaciones diferenciales como herramientas en modelos matemáticos, podemos estudiar problemas que surgen en disciplinas muy diversas. Desde sus comienzos han contribuido de manera destacada a solucionar muchos problemas y a interpretar numerosos fenómenos de la naturaleza. Su origen histórico está relacionado con sus aplicaciones a las ciencias naturales e ingeniería, ya que para resolver muchos problemas significativos se requiere la determinación de una función que debe satisfacer una ecuación en la que aparece su derivada.

Muchos fenómenos físicos pueden ser modelados matemáticamente por las ecuaciones diferenciales. Cuando la función que está siendo estudiada involucra dos o más variables independientes, la ecuación diferencial es por lo general una ecuación diferencial parcial. Dado que las funciones de varias variables son de cierta manera más complicadas que las de una variable, las ecuaciones diferenciales parciales puede llevar a algunos de los problemas más difíciles, pues la solución analítica puede ser difícil de encontrar o puede no existir y es ahí cuando se utilizan diversos métodos numéricos, para aproximar dicha solución.

El trabajo se centrará en saber aplicar estos métodos (diferencias finitas y elementos finitos) a ecuaciones de tipo hiperbólico (problemas que refieren fenómenos oscilatorios: vibraciones de cuerda, membranas, oscilaciones electromagnéticas), ecuaciones de tipo parabólico (problemas que se presentan al estudiar los procesos de conductibilidad térmica), ecuaciones de tipo elíptico (problemas que aparecen al estudiar procesos estacionarios, es decir, que no cambian con el tiempo) y a problemas de valor de frontera.

Las ecuaciones diferenciales parciales, se clasifican en lineales si la variable dependiente y todas sus derivadas aparecen elevadas a la primera potencia. Sea $u\,(x, y)$ una función de dos variables independientes, tenemos la forma general de una ecuación diferencial parcial de segundo grado

$$A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} + D\frac{\partial u}{\partial x} + E\frac{\partial u}{\partial y} + Fu = 0.$$

Y algunas de las ecuaciones en las cuales centraremos nuestra atención en nuestro estudio son: La ecuación de calor unidimensional

$$\frac{\partial u}{\partial t} = \alpha\frac{\partial^2 u}{\partial x^2},$$

la ecuación de onda:

$$\frac{\partial^2 u}{\partial t^2} = a^2\frac{\partial^2 u}{\partial x^2},$$

y la ecuación de Laplace en dos dimensiones:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Las cuales se pueden resolver por métodos analíticos como el de separación de variables donde se plantea una solución del tipo $u(x,t) = X(x)\,T(t)$, de manera que la ecuación diferencial parcial de segundo orden se transforma en dos ecuaciones diferenciales ordinarias, pero cuando esto no es posible es conveniente usar métodos numéricos, y sobre éstos centramos nuestra investigación.

Recalcando que el propósito final es proveer códigos de implementación para resolver problemas modelo de ecuaciones diferenciales parciales, por medio del método de diferencias finitas y por el método de elementos finito y estos códigos sean fácilmente adaptables por los usuarios para resolver cualquier otro problema de ecuaciones diferenciales parciales, dicho códigos son implementables en Octave y compatibles con MATLAB. Todos los métodos utilizados en el presente trabajo son convergentes.

3. Metodología

Se describe aquí los aspectos importantes de la metodología del presente trabajo de graduación:

1. Tipo de investigación. Este proyecto tiene las característica siguientes:

- *Bibliográfico*, porque se ha hecho una extensa recopilación de libros impresos y de libros obtenidos por Internet para contar con el suficiente material que cubra las necesidades del estudio. El objetivo es compilar coherentemente la información más útil y destacada del tema.
- *Descriptivo*, ya que se pretende estudiar a detalle la teoría preliminar y del tema en sí.
- *Sistemático*, porque los pseudocódigos implementados se diseñaran de manera eficiente para que se proprocione una solución aceptable.

2. Forma de Trabajo.

- Revisión de la bibliografía a utilizar.
- Se tendrán reuniones periódicas con el asesor del trabajo para tratar los diferentes aspectos de la investigación como estudiar y discutir la teoría y tratar los diferentes aspectos del trabajo escrito e implementación de algoritmos.
- Presentar mediante exposiciones los resultados.

3. Exposiciones.

Se tendrán dos exposiciones:

- Primera exposición: presentación del perfil del trabajo de graduación.
- Segunda exposición: presentación final del trabajo de graduación.

4. Capítulo 1. Preliminares

En este capítulo, presentamos la teoría básica para el estudio del tema que nos ocupa, vemos brevemente conceptos de ecuaciones diferenciales, álgebra lineal y análisis funcional.

4.1. Acerca de las ecuaciones diferenciales parciales.

4.1.1. ¿Qué es una ecuación diferencial parcial?

Una ecuación en derivadas parciales (EDP) de orden $n \in \mathbb{N}$ es una ecuación en la que aparece una función desconocida que depende (al menos) de dos variables reales, junto a algunas de sus derivadas parciales hasta de orden n. Cuando la función incógnita sólo depende de una variable real, se trata de una ecuación diferencial ordinaria (EDO) de orden n.

Se dice que una EDP es lineal si es lineal en la función desconocida y en todas sus derivadas parciales. En otro caso, se dice que es no lineal.

Dada una función $u(x, y)$, es habitual utilizar la siguiente notación abreviada para designar sus derivadas parciales

$$\frac{\partial u}{\partial x}(x,y) = u_x(x,y); \quad \frac{\partial u}{\partial y}(x,y) = u_y(x,y); \quad \frac{\partial^2 u}{\partial x^2}(x,y) = u_{xx}(x,y); \quad \frac{\partial^2 u}{\partial x \partial y}(x,y) = u_{yx}(x,y)$$

$$\frac{\partial^2 u}{\partial y \partial x}(x,y) = u_{xy}(x,y); \quad \frac{\partial^2 u}{\partial y^2}(x,y) = u_{yy}(x,y); ...$$

A partir de aquí, se supondrá que las funciones que manejamos son suficientemente regulares de forma que todas las derivadas parciales que aparecen están bien definidas y sean continuas.

Por otra parte, si la función u es de clase C^2 en un cierto dominio (existen todas las derivadas parciales hasta orden 2 de dicha función y son continuas en el dominio). Se sabe que $u_{yx}(x, y) = u_{xy}(x, y)$, gracias al Teorema de Schwarz (igualdad de las derivadas cruzadas). Por ello, en las EDP de segundo orden sólo aparecerá u_{xy} (y no $u_{yx}(x, y)$). En general, es irrelevante el orden en el cual se aplican k (ó menos) derivadas parciales a una función de clase C^k en un cierto dominio.

Definición (*Ecuación diferencial en derivadas parciales*) Se llama ecuación diferencial en derivadas parciales de orden m, a la ecuación de la forma:

$$F\left(x_1, x_2, ..., x_n, u, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, ..., \frac{\partial u}{\partial x_n}, ..., \frac{\partial^m u}{\partial^{k_1} x_1 \partial^{k_2} x_2 ... \partial^{k_n} x_n}\right) = 0, \tag{4.1}$$

que relaciona las variables independientes x_i, $\forall i = 1, 2, ..., n$, la función $u(x_1, x_2, ..., x_n)$ que se busca y sus derivadas parciales. Se cumple que k_i son enteros no negativos $\forall i = 1, 2, ..., n$, tales que $k_1 + k_2 + ... + k_n = m$.

Definición (*Orden*) Se llama orden de una EDP, al orden superior de las derivadas parciales que figuran en la ecuación.

Así, por ejemplo si x, y son variables independientes, $u = u(x, y)$ es la función buscada, entonces:

$$a) \quad yu_x - xu_y = 0 \qquad \text{es una EDP de 1er orden.}$$
$$b) \quad u_{xx} - u_{yy} = 0 \qquad \text{es una EDP de 2do orden.}$$

Definición (*Solución*) Sea la EDP definida en (4.1) de orden m, se llama solución de dicha EDP en cierta región D, a una función cualquiera $u = u(x_1, x_2, ..., x_n) \in C^m(D)$ (conjunto de funciones continuas en la región D que cuentan con todas las derivadas de hasta orden m inclusive), tal que al sustituir u, y sus derivadas en (4.1) la ultima se convierte en la identidad respecto a x_i, $\forall i = 1, 2, \cdots, n$ en la región D.

Hay tres ecuaciones diferenciales parciales básicas que son: la ecuación de onda, ecuación de calor y la ecuación de Laplace/Poisson. Todas las EDPs que estudiaremos a provienen de modelos físicos: la vibración vertical de las cuerdas de una guitarra o la membrana de un tambor; la evolución de la temperatura en piezas 1D, 2D o 3D; los equilibrios elásticos y térmicos de los problemas anteriores, etc. Esto proporciona una valiosa intuición del comportamiento que deben tener las soluciones de las EDPs consideradas y podremos interpretar físicamente los resultados obtenidos.

Podemos proponer una clasificación de las EDP's de segundo orden de dos variables independientes:

Definición: Sea la EDP de segundo orden

$$A(x,y)\frac{\partial^2 u(x,y)}{\partial x^2} + 2B(x,y)\frac{\partial^2 u(x,y)}{\partial x \partial y} + C(x,y)\frac{\partial^2 u(x,y)}{\partial y^2} + a(x,y)\frac{\partial u(x,y)}{\partial x} +$$

$$b(x,y)\frac{\partial u(x,y)}{\partial y} + c(x,y)u(x,y) = f(x,y),$$

en cierta región $\Omega \subset \mathbb{R}^2$. Se dice:

1. Hiperbólica en Ω si $\triangle = B^2 - AC > 0$ en Ω.
2. Parabólica en Ω si $\triangle = B^2 - AC = 0$ en Ω.
3. Elíptica en Ω si $\triangle = B^2 - AC < 0$ en Ω.

La ecuación del calor en 1D

Consideramos la evolución de la temperatura en una barra homogénea de longitud L sin focos ni sumideros de calor internos. Denotamos por $u(x, t)$ la temperatura del punto $x \in [0, L]$ en el instante $t \geq 0$. También podemos considerar una barra de longitud infinita, en cuyo caso $x \in \mathbb{R}$.

La EDP que modela la evolución de la temperatura es

$$u_t = k^2 u_{xx} \ , \ t \in (0, L) \ \ , \ t > 0.$$

El parámetro $k^2 = \kappa / (c\rho) > 0$ depende de la conductividad térmica κ, la densidad ρ y el calor específico c del material que conforma la barra.

La ecuación de ondas 1D (cuerda vibrante)

Consideramos el movimiento ondulatorio vertical de una cuerda vibrante horizontal de longitud L de densidad constante y composición homogénea no sometida a fuerzas externas. Denotamos por $u(x,t)$ el desplazamiento vertical respecto la posición de equilibrio del punto $x \in [0, L]$ de la cuerda en el instante $t \in \mathbb{R}$. También podemos considerar una cuerda vibrante de longitud infinita, en cuyo caso $x \in \mathbb{R}$. La EDP que modela el movimiento es

$$u_{tt} = c^2 u_{xx} \quad x \in (0, L) \quad t \in \mathbb{R}.$$

El parámetro $c > 0$ depende de las propiedades físicas del material y se interpreta como la velocidad a la que viajan las ondas en el material considerado.

Equilibrios elásticos y térmicos 1D

Equilibrio significa que el estado del cuerpo no cambia, sino que se mantiene estacionario en al tiempo. Se buscan soluciones $u = u(x)$ que no dependan del tiempo y así desaparecen las derivadas parciales u_t y u_{tt} . En tal caso, las EDPs $u_{tt} = c^2 u_{xx}$ y $u_t = k^2 u_{xx}$ se reducen a la EDO lineal de segundo orden $u'' = 0$, cuyas únicas soluciones son las funciones lineales de la forma $u(x) = ax + b$, con $a, b \in \mathbb{R}$.

Queda probado que los únicos equilibrios elásticos de una cuerda vibrante o equilibrios térmicos de una barra son los estados (desplazamiento o temperatura) lineales.

Las versiones multidimensionales

Antes de dar las versiones multidimensionales de las ecuaciones anteriores, se necesita generalizar el operador Laplaciano $\triangle$. Dada una función $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ que depende de n variables $\mathbf{x} = (x_1, ..., x_n)$, su Laplaciano es la suma de sus n derivadas parciales dobles:

$$\triangle u = \sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2} = u_{x_1 x_1} + ... + u_{x_n x_n}.$$

Por ejemplo, si la función u depende de una única variable x, entonces $\triangle u = u_{xx}$. En cambio, si depende de dos variables x, y, entonces $\triangle u = u_{xx} + u_{yy}$. Además, cuando la función dependa de la posición $\mathbf{x} = (x_1, ..., x_n)$ y el tiempo t, se interpreta que el Laplaciano sólo afecta a las variables de posición, sin incluir el término u_{tt} .

Las versiones n-dimensionales de las ecuaciones anteriores son las siguientes:

La ecuación de onda que modela el movimiento ondulatorio de un cuerpo elástico $\Omega \subset \mathbb{R}^n$ es

$$u_{tt} = c^2 \triangle u, \quad u = u(\mathbf{x}, t), \quad \mathbf{x} = (x_1, ..., x_n) \subset \Omega, \quad t \in \mathbb{R}.$$

La ecuación del calor que modela la evolución de la temperatura en un cuerpo $\Omega \subset \mathbb{R}^n$ es

$$u_t = k^2 \triangle u, \quad u = u(\mathbf{x}, t), \quad \mathbf{x} = (x_1, ..., x_n) \subset \Omega, \quad t > 0.$$

Desde un punto de vista físico, sólo interesan los casos 1D, 2D o 3D. Es decir, $n \leq 3$. Al igual que en las versiones 1D estamos suponiendo que el cuerpo es completamente homogéneo y que no existen fuerzas exteriores (ecuación de ondas) ni fuentes o sumideros de calor internas (ecuación de calor).

La ecuación de Laplace/Poisson

A partir de las versiones n-dimensionales de las ecuaciones de ondas y calor, vemos que tanto los equilibrios térmicos como los equilibrios elásticos de un cuerpo $\Omega \subset \mathbb{R}^n$ están modelados por la llamada ecuación de Laplace

$$\triangle u = 0, \quad u = u(\mathbf{x}, t), \quad \mathbf{x} = (x_1, ..., x_n) \subset \Omega.$$

La ecuación de Poisson es la versión no homogénea de la ecuación de Laplace. Consiste en, dada una función $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, buscar las soluciones de la ecuación

$$\triangle u = -F(x), \quad u = u(x, t), \quad x = (x_1, ..., x_n) \subset \Omega.$$

La ecuación de Poisson admite muchas interpretaciones físicas. Aquí tan sólo mencionamos que modela los equilibrios elásticos de un cuerpo $\Omega \subset \mathbb{R}^n$ sometido a la acción de una fuerza externa $F(x)$.

4.2. Condiciones iniciales y condiciones de frontera.

Cuando la ecuación estudiada tiene infinitas soluciones y queremos tener una solución concreta se añade a la ecuación un número adecuado de condiciones adicionales que pueden ser de dos tipos.

4.2.1. Condiciones iniciales: posición, velocidad y temperatura.

Estas condiciones fijan el estado del objeto en el instante inicial. Empezamos por la ecuación de onda, que es de segundo orden en el tiempo. Necesita exactamente dos condiciones iniciales:

La posición inicial : $u(x,0) = f(x)$ para $x \in \Omega$; y

La velocidad inicial : $u_t(x,0) = g(x)$ para $x \in \Omega$.

En cambio, la ecuación del calor es de primer orden en el tiempo:

La temperatura inicial : $u(x,0) = f(x)$ para $x \in \Omega$.

Y, la ecuación de Laplace es estática, por lo que no tiene sentido fijar el estado inicial del objeto, ya que ese estado es justamente la incógnita del problema.

4.2.2. Condiciones de frontera: Dirichlet (valor fijo) y Neumann (flujo fijo).

Estas condiciones (también llamadas condiciones de contorno) determinan la interacción del objeto con el medio que lo rodea, luego sólo tienen sentido cuando el objeto estudiado tiene frontera. Por ejemplo, en la ecuación de onda, la cuerda vibrante infinita no tiene frontera y las cuerdas de una guitarra sí. Hay de dos tipos de condiciones de frontera.

Tipo Dirichlet: Consisten en fijar el valor de la función incógnita en los puntos de la frontera.

Tipo Neumann: Consisten en "fijar el flujo" (es decir, el valor de la derivada en la dirección normal a la frontera) de la función incógnita en los puntos de la frontera.

Diremos que estas condiciones son homogéneas cuando el valor (o el flujo) fijado sea igual a cero.

Ejemplo 1. Un PVI de calor 1D en una barra de longitud L con condiciones de frontera de tipo Neumann consiste en las ecuaciones

$$\begin{cases} u_t = k^2 u_{xx} & x \in (0,L) \;\; t > 0, \\ u\left(x,0\right) = f\left(x\right) & x \in (0,L)\,, \\ u_x\left(0,t\right) = h_1\left(t\right) & t > 0, \\ u_x\left(L,t\right) = h_r\left(t\right) & t > 0, \end{cases}$$

donde la temperatura $f : [0,L] \to \mathbb{R}$ y los flujos h_1, $h_r : [0,+\infty) \to \mathbb{R}$ son funciones conocidas.

Ejemplo 2. Un problema de Poisson 2D en un cuadrado de lado $2L$ con condiciones de frontera de tipo Dirichlet homogéneas consiste en unas ecuaciones de la forma

$$\begin{cases} u_{xx} + u_{yy} = G\left(x,y\right) & x \in (-L,L) \;\; y \in (-L,L)\,, \\ u\left(\pm L, y\right) = 0 & y \in (-L,L)\,, \\ u\left(x, \pm L\right) = 0 & x \in (-L,L)\,. \end{cases}$$

4.2.3. Linealidad: superposición y homogeneización.

Existen varios trucos simples que se pueden aplicar en los problemas lineales, los cuales explicaremos a través de ejemplos concretos.

Superposición. Consideramos los dos PVIs de calor 1D en una barra de longitud L dados por

$$\begin{cases} v_t = k^2 v_{xx} & x \in (0,L) \ \ t > 0, \\ v(x,0) = f(x) & x \in (0,L), \\ v_x(0,t) = 0 & t > 0, \\ v_x(L,t) = 0 & t > 0. \end{cases} \qquad \begin{cases} w_t = k^2 w_{xx} & x \in (0,L) \ \ t > 0, \\ w(x,0) = 0 & x \in (0,L), \\ w_x(0,t) = h_1(t) & t > 0, \\ w_x(L,t) = h_r(t) & t > 0. \end{cases}$$

Ambos problemas tienen condiciones de frontera de tipo Neumann. La diferencia estriba en que el primero tiene una única condición no homogénea: la temperatura inicial, mientras que el segundo tiene dos: las condiciones de frontera en los extremos de la barra.

Entonces, dadas dos soluciones cualesquiera $v(x,t)$ y $w(x,t)$ de estos problemas, su superposición (suma) $u(x,t) = v(x,t) + w(x,t)$ es una solución del PVF de calor 1D presentado en el Ejemplo 1, que tiene tres condiciones no homogéneas.

En general, podemos "trocear" cualquier problema lineal en varios subproblemas de forma que cada subproblema tenga pocas (quizá incluso sólo una) ecuaciones/condiciones no homogéneas, siendo, por tanto, más simple que el problema original. En tal caso, si conseguimos resolver todos los subproblemas, la superposición (suma) de sus soluciones cumplirá el problema original.

Homogeneización. Este truco es similar al anterior, pero en vez de "trocear" el problema original en varios subproblemas simples, ahora queremos simplificarlo mediante un cambio de variables astuto.

En resumen, consideremos el PVF de calor 1D en una barra de longitud $L = 1$ con condiciones de frontera de tipo Dirichlet constantes

$$\begin{cases} u_t = k^2 u_{xx} & x \in (0,1) \ \ t > 0, \\ u(x,0) = x^2 & x \in (0,1), \\ u(0,t) = 1 & t > 0, \\ u(1,t) = 2 & t > 0. \end{cases}$$

La función $v(x) = x + 1$ cumple las condiciones de frontera: $v(0) = 1$ y $v(1) = 2$. Por tanto, si realizamos el cambio de variables $w(x,t) = u(x,t) - v(x)$, el problema original se transforma en

$$\begin{cases} w_t = k^2 w_{xx} & x \in (0,1) \ \ t > 0, \\ w(x,0) = x^2 - x - 1 & x \in (0,1), \\ w(0,t) = 0 & t > 0, \\ w(1,t) = 0 & t > 0, \end{cases}$$

que es un problema bastante más simple pues hemos homogeneizado las dos condiciones de frontera, sin deshomogeneizar la EDP.

4.3. Aproximando las derivadas

La determinación de la derivada de la función f en el punto x no es un problema numérico que sea tan fácil, específicamente, si $f(x)$ se puede calcular con n dígitos de precisión, es difícil calcular $f'(x)$ numéricamente con n dígitos de precisión. Esta dificultad se puede remontar a la resta entre cantidades que son casi iguales. En esta sección, estudiaremos varias alternativas para el cálculo numérico de $f'(x)$ y $f''(x)$.

4.3.1. Fórmula para la primera derivada vía series de Taylor

En primer lugar, se considera el método sobre la definición de $f(x)$. El cual consiste en seleccionar uno o más valores pequeños para h y escribimos:

$$f'(x) \approx \frac{1}{h}\left[f(x+h) - f(x)\right].$$

¿Cual es el error involucrado en esta fórmula?, para encontrarlo usamos el *Teorema de Taylor*, que se enuncia así: "Suponga que $f \in C^n[a,b]$, que $f^{(n+1)}$ existe en $[a,b]$ y $x_0 \in [a,b]$. Para cada $x \in [a,b]$, existe un número $\xi(x)$ entre x_0 y x tal que $f(x) = P_n(x) + R_n(x)$", donde

$$P_n = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2!}f''(x_0)(x-x_0)^2 + ... + \frac{1}{n!}f^{(n)}(x_0)(x-x_0)^n = \sum_{k=0}^{n}\frac{1}{k!}f^{(k)}(x_0)(x-x_0)^k$$

y

$$R_n = \frac{1}{(n+1)!}f^{(n+1)}(\xi(x))(x-x_0)^{n+1}.$$

En este caso, $P_n(x)$ es el n-ésimo polinomio de Taylor para f con respecto a x_0 y $R_n(x)$ se llama el termino del residuo o error de truncamiento asociado a $P_n(x)$. La serie infinita obtenida al tomar el límite de $P_n(x)$ cuando $n \to \infty$ es la serie de Taylor para f entorno a x_0.

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(\xi),$$

arreglado esta ecuación tenemos

$$f'(x) = \frac{1}{h}\left[f(x+h) - f(x)\right] - \frac{1}{2}hf''(\xi).$$

Entonces vemos que la aproximación dada tiene un error de $\frac{1}{2}hf''(\xi) = O(h)$, donde ξ esta en el intervalo que tiene por puntos finales x y $x+h$.

Esa última ecuación muestra que en general cuando $k \to 0$, la diferencia entre $f'(x)$ y la estimación $h^{-1}\left[f(x+h) - f(x)\right]$ se aproxima a cero con la misma tasa que h lo hace, es decir $O(h)$. De hecho, si $f''(x) = 0$, entonces el termino del error seria $\frac{1}{6}h^2 f'''(\gamma)$, el cual converge a cero con la misma rapidez que $O(h^2)$, pero usualmente $f''(x)$ no es cero.

Es ventajoso tener una convergencia en los procesos numéricos que se aproxime a cero con potencias superiores, para el caso queremos aproximar $f'(x)$ con un error que se comporte como $O(h^2)$, para obtenerlo desarrollamos las dos series de Taylor

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2!}h^2 f''(x) + \frac{1}{3!}h^3 f'''(x) + \frac{1}{4!}h^4 f^{(4)}(x) + ... \quad (4.2)$$

$$f(x-h) = f(x) - hf'(x) + \frac{1}{2!}h^2 f''(x) - \frac{1}{3!}h^3 f'''(x) + \frac{1}{4!}h^4 f^{(4)}(x) - ... \quad (4.3)$$

y restando tenemos

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{2}{3!}h^3 f'''(x) + \frac{2}{5!}h^5 f^{(5)}(x) + ...$$

Esto conduce a una fórmula muy importante para aproximar $f'(x)$:

$$f'(x) = \frac{1}{2h}\left[f(x+h) - f(x-h)\right] - \frac{1}{3!}h^2 f'''(x) - \frac{1}{5!}h^4 f^{(5)}(x) - ...$$

Expresando de otro modo

$$f'(x) \approx \frac{1}{2h}\left[f(x+h) - f(x-h)\right],$$

con un término para el error de $-\frac{1}{3!}h^2 f'''(x)$ que hace que sea $O(h^2)$.

4.3.2. Fórmulas para la segunda derivada a través de series de Taylor

En la solución numérica de las ecuaciones diferenciales, a menudo es necesario aproximar segundas derivadas. Vamos a deducir la fórmula más importante para lograr esto. Basta con sumar las dos series de Taylor de la ecuación $(4.2), (4.3)$ y se obtiene

$$f(x+h) + f(x-h) = 2f(x) + h^2 f''(x) + 2\left[\frac{1}{4!}h^4 f^{(4)}(x) + ...\right]$$

cuando despejamos $f''(x)$ obtenemos

$$f''(x) = \frac{1}{h^2}\left[f(x+h) - 2f(x) + f(x-h)\right] - \frac{2}{4!}h^2 f^{(4)}(x),$$

y llevando a cabo el mismo proceso usando la fórmula de Taylor con resto, se obtiene que $E = -\frac{1}{12}h^2 f^{(4)}(\xi)$ para algún ξ en el intervalo $(x-h, x+h)$. Y hemos obtenido la aproximación

$$f''(x) \approx \frac{1}{h^2}\left[f(x+h) - 2f(x) + f(x-h)\right],$$

con error $O(h^2)$.

4.4. Definiciones en álgebra lineal

Definición. Se dice que una matriz A de $n \times n$ es invertible, si existe una matriz A^{-1} de $n \times n$, con $AA^{-1} = A^{-1}A = I$, donde I es la matriz identidad. La matriz A^{-1} se llama inversa de A. Una matriz que no tiene inversa se le da el nombre de no invertible o singular.

Definición. Se dice que una matriz A de $n \times n$ es diagonal dominante estricta cuando

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|,$$

es aplicable para toda $i = 1, 2, ..., n$.

Definición. Una matriz A es definida positiva si es simétrica y si $\mathbf{x}^t A\mathbf{x} > 0$ para todo vector columna n dimensional $\mathbf{x} \neq 0$.

Definición. (Matriz tridiagonal) Una matriz $A = (a_{ij})$ de $n \times n$ se dice tridiagonal si $a_{ij} = 0$, siempre que $|i - j| > 1$, esto es

$$\begin{pmatrix} a_{11} & a_{12} & 0 & \dots & \dots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & a_{n-1,1} \\ 0 & \dots & \dots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}$$

4.4.1. Sistemas tridiagonales y banda

En muchas aplicaciones, se encuentran sistemas lineales extremadamente grandes que tienen estructura de banda. Las matrices banda a menudo ocurren en la solución de ecuaciones diferenciales ordinarias y parciales. Es una ventaja desarrollar códigos computacionales diseñados específicamente para tales sistemas lineales, ya que reducen la cantidad de espacio utilizado.

En la práctica es importante el sistema tridiagonal. Aquí, todos los elementos distintos de cero en la matriz de coeficientes deben estar en la diagonal principal o en las dos diagonales justo por encima y por debajo de la diagonal principal (por lo general llamada superdiagonal y subdiagonal, respectivamente):

$$\begin{bmatrix} d_1 & c_1 & & & & & & \\ a_1 & d_2 & c_2 & & & & & \\ & a_2 & d_3 & c_3 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & a_{i-1} & d_i & c_i & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & a_{n-2} & d_{n-1} & c_{n-1} \\ & & & & & & a_{n-1} & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_i \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \tag{4.4}$$

Todos los elementos que no están en las diagonales mostradas son 0´s . Una matriz tridiagonal es caracterizada por la condición $a_{ij} = 0$ si $|i - j| \geq 2$. En general se dice que una matriz tiene estructura de banda si existe un entero k (menor que n) tal que $a_{ij} = 0$ siempre que $|i - j| \geq k$.

Los requisitos de almacenamiento de una matriz banda son menores que los de una matriz en general, del mismo tamaño. Entonces, una matriz diagonal $n \times n$ requiere sólo n localizaciones en la memoria de una computadora, y una matriz tridiagonal requiere sólo $3n - 2$. Este hecho es importante si se utilizan matrices banda de orden muy grande.

Para matrices banda, el algoritmo de eliminación gaussiana podría hacerse muy eficiente si se sabe de antemano que el pivoteo es innecesario, a continuación se desarrollará un código para resolver un sistema tridiagonal.

Escribimos un programa diseñado para resolver un sistema lineal de n ecuaciones con n incógnitas, como el que se muestra en la ecuación (2.4) .

Tanto la fase de eliminación hacia adelante y la fase de sustitución hacia atrás se incorporan en el procedimiento, y ningún pivote es usado, es decir, las ecuaciones de pivote son los dados por el ordenamiento natural $\{1, 2, ..., n\}$. Por lo tanto, se utiliza eliminación de Gauss de forma trivial.

En el paso 1, restamos a_1/d_1 veces la fila 1 a la fila 2, creando así un 0 en la posición a_1. Sólo las entradas d_2 y b_2 son alteradas. Observe que c_2 no se altera.

En el paso 2, se repite el proceso, utilizando la nueva fila 2 como la fila pivote. Así es como el d_i´s y de b_i's se modifican en cada paso:

$$\begin{cases} d_2 \leftarrow d_2 - \left(\frac{a_1}{d_1}\right) c_1 \\ b_2 \leftarrow b_2 - \left(\frac{a_1}{d_1}\right) b_1 \end{cases}$$

y en general, obtenemos

$$\begin{cases} d_i \leftarrow d_i - \left(\frac{a_{i-1}}{d_{i-1}}\right) c_{i-1} \\ b_i \leftarrow b_i - \left(\frac{a_{i-1}}{d_{i-1}}\right) b_{i-1} \end{cases} \quad (2 \leq i \leq n)$$

Al final de la fase de eliminación hacia adelante, la forma del sistema es el siguiente:

$$\begin{bmatrix} d_1 & c_1 & & & & & & \\ & d_2 & c_2 & & & & & \\ & & d_3 & c_3 & & & & \\ & & & \ddots & \ddots & & & \\ & & & & d_i & c_i & & \\ & & & & & \ddots & \ddots & \\ & & & & & & d_{n-1} & c_{n-1} \\ & & & & & & & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_i \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}$$

Por supuesto, que los b_i's y d_i' s no son lo que eran en el comienzo de este proceso, pero los c_i's si lo son.

La fase de sustitución hacia atrás resuelve de $x_n, x_{n-1}, ..., x_1$ como sigue:

$$\begin{aligned} x_n &\leftarrow \frac{b_n}{d_n}, \\ x_{n-1} &\leftarrow \frac{1}{d_{n-1}} \left(b_{n-1} - c_{n-1} x_n \right), \end{aligned}$$

Finalmente, obtenemos $x_i \leftarrow \frac{1}{d_i} \left(b_i - c_i x_{i+1} \right)$ para $i = n-1, n-2, ..., 1$.

Escribimos el pseudocódigo para resolver un sistema tridiagonal, utilizamos vectores (a_i), (d_i), y (c_i) de las diagonales de la matriz de coeficientes y el vector (b_i) para el lado derecho, y para que guarde la solución el vector (x_i). Note que se estan modificado los datos originales en los vectores (d_i) y (b_i). En MATLAB podemos adecuar este procedimiento para que sea una función (que llamaremos tri) que nos devuelva la solución (x_i), así

```
% metodo de gauss pasa sistemas tridiagonales,
function x= tri(a,d,c,b)
% INPUT: a,d,c,b vectores de longitud n
%OUTPUT: x vector solucion del sistema tridiagonal A*x=b
% donde A(1:n,1:n)=d, A(1:n-1,2:n)=a, A(2:n,1:n-1)=c;

%Ejemplo: Resolver [1 4 0;2 -1 1;0 -2 1][x y z]=[17 9 4]
% a=[2 2 0], d=[1 -1 -1], c=[4 1 0], b=[17 9 0]

n1 = length(a); n2 = length(b); n3 = length(c); n4 = length(d);
if n1~=n2 || n1~=n3 || n1~=n4
   error('a,b,c y d tienen que tener la misma longitud.')
end

%longitud común
n=n1;
for i=2:n
   xmult=a(i-1)/d(i-1);
   d(i)=d(i)-xmult*c(i-1);
   b(i)=b(i)-xmult*b(i-1);
end
x(n)=b(n)/d(n);
for i=n-1:-1:1;
   x(i)=(b(i)-c(i)*x(i+1))/d(i);
end
end
```

Por ejemplo: Resolver el sistema tridiagonal,

$$\begin{array}{rrrcr} x_1 & +4x_2 & & = & 17 \\ 2x_1 & -x_2 & +x_3 & = & 9 \\ & 2x_2 & -x_3 & = & 4 \end{array}$$

Los vectores para nuestra función son: $\mathbf{a} = [2, 2, 0]$, $\mathbf{d} = [1, -1, -1]$, $\mathbf{c} = [4, 1, 0]$ y $\mathbf{b} = [17, 9, 4]$ y la solución es $\mathbf{x} = [5, 3, 2]$.

Los sistemas tridiagonales aparecen en problemas con ecuaciones diferenciales parciales aplicando el método de Crank-Nicolson para encontrar la solución, por lo que se usará bastante esta función *tri* en MATLAB y Octave.

Veamos la definición formal de una matriz banda:

Definición. Una matriz de $n \times n$ recibe el nombre de matriz banda si existen los enteros p y q con $1 < p, q < n$, que tienen la propiedad de que $a_{ij} = 0$ siempre que $i + p \leq j$ o $j + q \leq i$. El ancho de la banda de estas matrices se define como $w = p + q - 1$.

Por ejemplo la matriz $A = \begin{bmatrix} 7 & 2 & 0 \\ 2 & 5 & -1 \\ 0 & -5 & -6 \end{bmatrix}$ es una matriz banda con $p = q = 2$ y con ancho de banda $2 + 2 - 1 = 3$.

La definición de matriz banda hace que estas matrices concentren todos sus elementos distintos de cero alrededor de la diagonal. Dos casos especiales de matrices banda que ocurren a menudo en la práctica tienen $p = q = 2$ y $p = q = 4$.

Las matrices de ancho de banda 3, que se presentan cuando $p = q = 2$, como ya lo hemos visto se llaman tridiagonales por tener la forma

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & \cdots & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & & & \vdots \\ 0 & q_{32} & a_{33} & a_{34} & \ddots & & \vdots \\ \vdots & \ddots & & \ddots & & \ddots & 0 \\ \vdots & & \ddots & & \ddots & & a_{n-1,n} \\ 0 & \cdots & \cdots & \cdots & 0 & a_{n,n-1} & a_{nn} \end{bmatrix}$$

Los algoritmos de factorización o de resolución de un sistema de ecuaciones lineales pueden simplificarse considerablemente en el caso de las matrices banda, porque una gran cantidad de ceros aparecen en ellas en patrones regulares. Es interesante señalar la forma que en este caso asume el método de Crout.

Recordamos que el método de Crout es aplicable a una matriz cuadrada A y se obtiene una factorización de tipo $A = LU$, donde L es una matriz triangular inferior y U es una matriz triangular superior cuyos elementos de la diagonal son todos 1.

Para ilustrar lo anterior en el caso de factorización, supóngamos que podemos factorizar una matriz tridiagonal A en matrices tridiagonales inferior L y superior U. Puesto que A tiene sólo $(3n-2)$ elementos distintos de cero, habrá apenas $(3n-2)$ condiciones aplicables para determinar los elementos de L y U, naturalmente a condición de que tambien se obtengan los elementos ceros de A. Supongamos que queremos encontrar las matrices en la forma

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & \cdots & 0 \\ l_{21} & l_{22} & \ddots & & \vdots \\ 0 & & & \ddots & \vdots \\ \vdots & \ddots & & & 0 \\ 0 & \cdots & 0 & l_{n,n-1} & l_{nn} \end{bmatrix} \text{ y } U = \begin{bmatrix} 1 & u_{12} & 0 & \cdots & 0 \\ 0 & & & \ddots & \vdots \\ \vdots & \ddots & & & 0 \\ \vdots & & \ddots & & u_{n-1,n} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

Hay $(2n-1)$ elementos determinados de L y $(n-1)$ elementos indeterminados de U, que suman el número total de condiciones $(3n-2)$. Los elementos cero de A se obtienen automáticamente. La multiplicación que incluye $A = LU$ nos da, además de los elementos cero,

$$a_{11} = l_{11}; \tag{4.5}$$

$$a_{i,i-1} = l_{i,i-1} \text{ para cada } i = 2, 3, ..., n; \tag{4.6}$$

$$a_{ii} = l_{i,i-1}u_{i-1,i} + l_{ii}, \text{ para cada } i = 2, 3, ..., n; \tag{4.7}$$

$$a_{i,i+1} = l_{ii}u_{i,i+1}, \text{ para cada } i = 1, 2, ..., n-1 \tag{4.8}$$

Una solución de este sistema se obtiene aplicando primero la ecuación (4.6) para obtener el término fuera de la diagonal L y luego la ecuaciones (4.7) y (4.8) para obtener alternativamente el resto de elementos de U y de L. Éstos pueden guardarse en los elementos correspondientes de A.

El programa que se verá en la siguiente sección, está escrito en GNU Octave (compatible con MATLAB), el cuál resuelve un sistema de ecuaciones lineales de $n \times n$ cuya matriz de coeficientes es tridiagonal. Sólo requiere $(5n-4)$ multiplicaciones/divisiones y $(3n-3)$ sumas/restas. En consecuencia, ofrece una importante ventaja computacional sobre los métodos que no cuentan con la matriz tridiagonal.

4.4.2. Factorización de Crout de sistemas lineales tridiagonales

Para resolver el sistema lineal $n \times n$:

$$\begin{array}{rrrrcl}
a_{11}x_1 & +a_{12}x_2 & & & = & b_1 \\
a_{21}x_1 & +a_{22}x_2 & +a_{23}x_3 & & = & b_2 \\
 & & \vdots & & \vdots & \\
 & a_{n-1,n-2}x_{n-2} & +a_{n-1,n-1}x_{n-1} & +a_{n-1,n}x_n & = & b_{n-1} \\
 & & a_{n,n-1}x_{n-1} & +a_{nn}x_n & = & b_n
\end{array}$$

que se supone tiene solución única.

ENTRADA Los elementos de la matriz A y el vector b

SALIDA La solución $x_1, x_2, ..., x_n$ elementos en el vector x.

```
% metodo de Crout,
function x= crout(A,b)
%A es la matriz de coeficientes y b el vector respuesta
%vector x solucion del sistema tridiagonal A*x=b
%Ejemplo: resolver
%[2 -1 0 0;-1 2 -1 0;0 -1 2 -1;0 0 -1 2][x;y;z;w]=[1;0;0;1]
%A= [2 -1 0 0;-1 2 -1 0;0 -1 2 -1;0 0 -1 2], b=[1 0 0 1]
n = length(b);
l = zeros(n);
u = zeros(n);
z = zeros(1,n);

% resolvemos Lz=b
l(1,1)=A(1,1);
u(1,2)=A(1,2)/l(1,1);
z(1)=b(1)/l(1,1);

for i= 2:n-1
    l(i,i-1)=A(i,i-1);
    l(i,i)=A(i,i)-l(i,i-1)*u(i-1,i);
    u(i,i+1)=A(i,i+1)/l(i,i);
    z(i)=(b(i)-l(i,i-1)*z(i-1))/l(i,i);
end

l(n,n-1)=A(n,n-1);
l(n,n)=A(n,n)-l(n,n-1)*u(n-1,n);
z(n)=(b(n)-l(n,n-1)*z(n-1))/l(n,n);

%resuelve Ux=z
x(n)=z(n);
for i=n-1:-1:1
    x(i)=z(i)-u(i,i+1)*x(i+1);
end
end
```

Ejemplo: Para explicar con un ejemplo el procedimiento de las matrices tridiagonales, consideramos el sistema tridiagonal de ecuaciones

$$\begin{array}{rrrrcl}
2x_1 & -x_2 & & & = & 1 \\
-x_1 & +2x_2 & -x_3 & & = & 0 \\
 & -x_2 & +2x_3 & -x_4 & = & 0 \\
 & & -x_3 & +2x_4 & = & 1
\end{array}$$

cuya matriz aumentada es

$$\left[\begin{array}{cccc:c} 2 & -1 & 0 & 0 & 1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & 1 \end{array}\right]$$

El algoritmo de factorización de Crout genera la factorización

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & \frac{3}{2} & 0 & 0 \\ 0 & -1 & \frac{4}{3} & 0 \\ 0 & 0 & -1 & \frac{5}{4} \end{bmatrix} \begin{bmatrix} 1 & \frac{-1}{2} & 0 & 0 \\ 0 & 1 & \frac{-2}{3} & 0 \\ 0 & 0 & 1 & \frac{-3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix} = LU$$

Al resolver el sistema $Lz = b$ obtenemos $z = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, 1\right)^t$ y la solución de $Ux = z$ es $x = (1, 1, 1, 1)^t$.

El algoritmo de factorización de Crout puede aplicarse siempre que $l_{ii} \neq 0$ para toda $i = 1, 2, ..., n$. Dos condiciones que garantizan la veracidad de esto son que la matrices de coeficientes del sistema sea definida positiva o que sea diagonal dominante estricta. Una condición adicional con que se garantiza la aplicabilidad del algoritmo se da en el siguiente teorema.

Teorema. Supongamos que $A = (a_{ij})$ es tridiagonal con $a_{i,i-1}a_{i,i+1} \neq 0$ para toda $i = 2, 3, ..., n-1$. Si $|a_{11}| > |a_{12}|$, $|a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|$para cada $i = 2, 3, ..., n-1$ y $|a_{nn}| > |a_{n,n-1}|$ entonces A es no singular y los valores de l_{ii} descritos en el algoritmo de factorización de Crout son distintos de cero para cada $i = 1, 2, ..., n$.

Demostración.

Veamos primero que los elementos l_{ii} son todos distintos de cero:

Entonces $|l_{11}| = |a_{11}| > 0$ y $|u_{12}| = \frac{|a_{12}|}{|l_{11}|} < 1$. En general asumimos que $|l_{jj}| > 0$ y $|u_{j,j+1}| < 1$, para $j = 1, ..., i-1$. Entonces,

$$|l_{ii}| = |a_{ii} - l_{i,i-1}u_{i-1,i}| = |a_{ii} - a_{i,i-1}u_{i-1,i}| \geq |a_{ii}| - |a_{i,i-1}u_{i-1,i}| > |a_{ii}| - |a_{i,i-1}| > 0,$$

y

$$|u_{i,i+1}| = \frac{|a_{i,i+1}|}{|l_{ii}|} < \frac{|a_{i,i+1}|}{|a_{ii}| - |a_{i,i-1}|} \leq 1,$$

para $i = 2, ..., n-1$. Además

$$|l_{nn}| = |a_{nn} - l_{n,n-1}u_{n-1,n}| = |a_{nn} - a_{n,n-1}u_{n-1,n}| \geq |a_{nn}| - |a_{n,n-1}| > 0,$$

así tenemos que para el determinante de A,

$$\det A = \det L \cdot \det U = l_{11} \cdot l_{22} ... l_{nn} \cdot 1 > 0,$$

por lo que A es invertible (no singular).

4.5. Definiciones en análisis funcional

En esta sección introducimos brevemente algunos de los conceptos que necesitaremos en el estudio del método de elementos finitos.

4.5.1. Norma y seminorma.

Consideremos un cuerpo F(en general, $\mathbb{R}$ o $\mathbb{C}$), y tomemos sobre él un F-espacio vectorial E.

Definición (norma) Una norma es una función sobre el espacio vectorial, que usualmente se denota $\|.\|_E : E \to \mathbb{R}_0^+$ que tiene las propiedades:

I.$\|\mathbf{x}+\mathbf{y}\|_E \leq \|\mathbf{x}\|_E + \|\mathbf{y}\|_E \quad \forall\, \mathbf{x}, \mathbf{y} \in E$,

II.$\|\lambda \mathbf{x}\|_E = |\lambda| \, \|\mathbf{x}\|_E \quad \forall\, \lambda \in F$,

III.$\|\mathbf{x}\|_E = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0}$.

Definición (seminorma) Se define una seminorma como una función sobre el espacio vectorial, a valores en el cuerpo, de manera que valen las propiedades I y II de una norma. Está claro que toda norma es una seminorma.

Definición (espacio normado) Un espacio normado es un par $(E, \|.\|_E)$ formado por un espacio vectorial E sobre un cuerpo F, y una norma a valores en el cuerpo F.

Si miramos con cuidado las propiedades I, II y III de la definición de norma; podemos notar que todo espacio normado es un espacio vectorial métrico, donde la métrica tiene las propiedades adicionales:

a) $d(\alpha x, \alpha y) = |\alpha| \, d(x, y)$,

b) $d(x + z, y + z) = d(x, y)$.

Podemos entonces hablar de continuidad, y como en el caso de un espacio métrico, es trivial la verificación de que la función $\|.\| : E \to \mathbb{R}^+$ es una función continua, simplemente reescribiendo la desigualdad triangular (propiedad I)$|\|x\| + \|y\|| \leq \|x + y\|$,

Además las desigualdades

$$\|x_1 + y_1 - (x_2 + y_2)\| \leq \|x_1 - x_2\| + \|y_1 - y_2\|,$$

y

$$\begin{aligned} \|\lambda_1 x_1 - \lambda_2 x_2\| &\leq \|\lambda_1 x_1 - \lambda_1 x_2\| + \|\lambda_1 x_2 - \lambda_2 x_2\| \\ &= |\lambda| \, \|x_1 - x_2\| + |\lambda_1 - \lambda_2| \, \|x_2\| \, , \end{aligned}$$

prueban que tanto la suma como el producto por escalares son funciones continuas en cualquier espacio normado.

4.5.2. Producto interno

Definición. Sea X un espacio vectorial sobre K (donde K es $\mathbb{R}$ o $\mathbb{C}$). Un producto interno en X es una función tal que

(a) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ para todo $x, y \in X$.

(b) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ para todo $x, y, z \in X$.

(c) $\langle \lambda x, y\rangle = \lambda \langle x, y\rangle$ para todo $x, y \in X$, para todo $\lambda \in K$.

(d) $\langle x, x\rangle \geq 0$ para todo $x \in X$.

(e) $\langle x, x\rangle = 0$ si y sólo si $x = 0$.

Además decimos que $(X, \langle ., .\rangle)$ es un espacio con producto interno .

Ejemplo Sea $\mathbb{C}^n$ el espacio de los vectores $1 \times n$. $\mathbb{C}^n$ es un espacio con producto interno con el producto escalar euclídeo

$$\langle v, w\rangle_n = vw^* = \sum_{k=1}^{n} v_k \overline{w_k},$$

donde $v, w \in \mathbb{C}^n$, $v = (v1, ..., vn)$, $w = (w1, ..., wn)$, y w^* es el adjunto de w.

4.5.3. Espacios de Hilbert

Un espacio de Hilbert es una generalización del concepto de espacio euclídeo. Esta generalización permite que nociones y técnicas algebraicas y geométricas aplicables a espacios de dimensión dos y tres se extiendan a espacios de dimensión arbitraria, incluyendo a espacios de dimensión infinita. Ejemplos de tales nociones y técnicas son la de ángulo entre vectores, ortogonalidad de vectores, el teorema de Pitágoras, proyección ortogonal, distancia entre vectores y convergencia de una sucesión.

Más formalmente, se define como un espacio de producto interior que es completo con respecto a la norma vectorial definida por el producto interior.

Cada producto interior $\langle ., .\rangle$ en un espacio vectorial H, que puede ser real o complejo, da lugar a una norma $\|.\|$ que se define como sigue:

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

H es un espacio de Hilbert si es completo con respecto a esta norma. Completo en este contexto significa que cualquier sucesión de Cauchy de elementos del espacio converge a un elemento en el espacio, en el sentido que la norma de las diferencias tiende a cero.

Se dan otras dos definiciones de espacios, que nos serán útiles,

Un espacio de Sóbolev es un tipo de espacio vectorial funcional, dotado de una norma de tipo L^p, tal que la función y sus derivadas hasta cierto orden tienen norma finita. Un espacio de Sóbolev puede ser considerado como un subespacio de un espacio L^p.

Un espacio de Sóbolev es un espacio vectorial normado de funciones que puede verse como un subespacio de un espacio L^p. De hecho un espacio de Sóbolev es un subespacio vectorial del espacio L^p formado por clases de funciones tales que sus derivadas hasta orden m pertenecen también a L^p. Dado un dominio $\Omega \subset \mathbb{R}^n$ el espacio de Sobolev $W^{m,p}(\Omega)$, se define como:

$$W^{m,p}(\Omega) = \{ f \in L^p(\Omega) |\ D^\alpha f \in L^p(\Omega),\ \forall \alpha \in \mathbb{N}^n : |\alpha| \leq m \ \} \subset L^p(\Omega),$$

donde $D^\alpha f$, es la notación multi-índice para las derivadas parciales. Debe tenerse presente que dicho espacio está de hecho formado realmente por clases de equivalencia de funciones.

Un espacio de Banach es un espacio vectorial normado completo. Esto quiere decir que un espacio de Banach es un espacio vectorial V sobre el cuerpo de los números reales o el de los complejos con una norma $\|\cdot\|$, tal que toda sucesión de Cauchy en V tiene un límite en V.

4.5.4. Distribuciones

Una distribución o función generalizada es un objeto matemático que generaliza la noción de función y la de medida. La noción de distribución sirve para extender el concepto de derivada a todas las funciones localmente integrables y a entes aún más generales.

Una distribución convencional sobre Ω es un elemento del espacio dual topológico del espacio vectorial de funciones de clase $C^\infty(\Omega)$ sobre un cierto conjunto $\Omega \subset \mathbb{R}^n$ cuyo soporte es un conjunto compacto. Es decir, una distribución es una función lineal y continua definida sobre un cierto espacio de funciones diferenciables definidas sobre conjuntos cerrados contenidos en Ω. Las funciones definidas sobre el conjunto Ω se llama espacio de funciones test.

Deben ser infinitamente diferenciables, es decir de clase, C^∞. Deben ser funciones cuyo soporte sea compacto.

Soporte compacto

Se dice que una función test ϕ tiene soporte compacto si el conjunto de puntos

$$K = \overline{\{x/\phi(x) \neq 0\}},$$

donde la función es diferente de cero es compacto. Se dice que una distribución S tiene soporte compacto si existe un conjunto compacto K de U tal que para cada función test ϕ cuyo soporte no se interseca con K se tiene que $S(\phi) = 0$. Alternativamente podemos definir las distribuciones de soporte compacto como funciones lineales continuas sobre el espacio $C^\infty(U)$, con una topología definida sobre este espacio por la convergencia uniforme.

Derivada de una distribución

El concepto de derivada distribucional o derivada en el sentido de las distribuciones generaliza el concepto de derivada ordinaria a distribuciones y funciones no continuas. Esta extensión se realiza a partir del procedimiento de integración por partes. Dada una distribución o función discontinua f; su derivada en el sentido de las distribuciones se define simplemente como la única función f'; que satisface $\forall\phi$

$$\int_\Omega f'\phi = -\int_\Omega f\phi'.$$

Algunos ejemplos de derivadas en el sentido distribucional son:

La derivada en el sentido de las distribuciones de una función diferenciable coincide con su derivada ordinaria.

La función valor absoluto tiene por derivada distribucional la función signo.

Más prácticamente: Dada $u \in L^2(I)$, si $\exists g \in L^2(I)$ y $\forall\varphi \in D(I)$ se cumple que

$$-\int_I u\varphi' = \int_I g\varphi,$$

entonces decimos que $g = u'$ en el sentido de las distribuciones.

5. Capítulo 2. Método de diferencias finitas

El método de diferencias finitas es un método universal para resolver ecuaciones diferenciales. En este capítulo, para una ecuación diferencial parcial parabólica, presentamos algunos esquemas de diferencias y analizarmos su convergencia. Presentamos también el teorema de la Equivalencia de Lax y sus resultados teóricos relacionados.

El método de diferencias finitas puede ser difícil de analizar, en parte porque es bastante general en su aplicabilidad. Gran parte de la existencia de estabilidad y el análisis de convergencia se limitan a casos especiales, en particular para ecuaciones diferenciales lineales con coeficientes constantes. Estos resultados se usan entonces para predecir el comportamiento de los métodos de diferencias para ecuaciones más complicadas.

5.1. Aproximaciones en diferencias finitas

La idea básica del método de diferencias finitas es aproximar las diferenciales por diferencias apropiadas, reduciendo así una ecuación diferencial a un sistema algebraico. Hay muchas maneras de hacer la aproximación.

Supongamos que f es una función real diferenciable en $\mathbb{R}$. Sea $x \in \mathbb{R}$ y $h > 0$. Entonces tenemos los siguientes tres aproximaciones populares en diferencias:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}, \tag{5.1}$$

$$\frac{f(x) - f(x-h)}{h}, \tag{5.2}$$

$$\frac{f(x+h) - f(x-h)}{2h}. \tag{5.3}$$

Estas diferencias se llaman diferencias hacia adelante, hacia atrás y diferencias centradas, respectivamente tal como vimos en la sección anterior. Suponiendo que f tiene segunda derivada, se comprueba que los errores de aproximación para el avance y retroceso diferencias son ambos $O(h)$. Si existe la tercera derivada de f, entonces el error de aproximación para la diferencia centrada es $O(h^2)$. Vemos que si la función es suave, la diferencia centrada es una aproximación más precisa en la derivada.

La segunda derivada de la función f se aproxima por lo general con diferencias centradas de segundo orden:

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{5.4}$$

Se ha verificado antes que cuando f tiene un cuarta derivada, la aproximación de error es $O(h^2)$. Ahora vamos a utilizar estas fórmulas de diferencias para formular algunos esquemas en diferencias de un problema con valores iniciales y de contorno como ejemplo para la ecuación

de calor.

Ejemplo 5.1 Consideremos el problema

$$\begin{aligned} u_t &= \nu u_{xx} + f(x,t) \quad \text{en } (0,\pi)\times(0,T), && (5.5)\\ u(0,t) &= u(\pi,t) = 0, \quad 0 \le t \le T, && (5.6)\\ u(x,0) &= u_0(x), \quad 0 \le x \le \pi. && (5.7) \end{aligned}$$

La ecuación diferencial (5.5) se puede utilizar para modelar una variedad de procesos físicos tales como la conducción de calor. Aquí $\nu > 0$ es una constante, f y u_0 son funciones continuas. Para desarrollar un método de diferencias finitas, necesitamos introducir una malla de puntos. Sean N_x y N_t enteros positivos, $h_x = \pi/N_x$, $h_t = T/N_t$ y definimos los puntos de la partición

$$\begin{aligned} x_j &= jh_x, \quad j = 0,1,...,N_x,\\ t_m &= mh_t, \quad m = 0,1,...,N_t. \end{aligned}$$

Un punto de la forma (x_j, t_m) se llama un punto malla y nos interesa calcular los valores de la solución aproximada en los puntos de malla. Usamos la notación v_j^m para una aproximación a $u_j^m \equiv u(x_j,t_m)$ calculada a partir de un esquema de diferencias finitas. Escribimos $f_j^m = f(x_j,t_m)$ y

$$r = \nu h_t/h_x^2.$$

Entonces podemos escribir varios esquemas de diferencias finitas.

El primero es

$$\begin{aligned} \frac{v_j^{m+1}-v_j^m}{h_t} &= \nu\frac{v_{j+1}^m - 2v_j^m + v_{j-1}^m}{h_x^2} + f_j^m, 1 \le j \le N_x - 1,\ 0 \le m \le N_t - 1, && (5.8)\\ v_0^m &= v_{N_x}^m = 0, \ \ 1 \le m \le N_t, && (5.9)\\ v_j^0 &= u_0(x_j), \ \ 0 \le j \le N_x. && (5.10) \end{aligned}$$

Este esquema se obtiene por la discretización de la ecuación diferencial (5.5) en $x = x_j$ y $t = t_m$, sustituyendo la derivada en el tiempo con una diferencia hacia adelante y la segunda derivada espacial con una diferencia centrada de segundo orden. Por lo tanto, se denomina un esquema de espacio-tiempo centrado hacia adelante. La ecuación en diferencias (5.8) se puede escribir como

$$v_j^{m+1} = (1-2r)v_j^m + r\left(v_{j+1}^m + v_{j-1}^m\right) + h_t f_j^m, \ \ 1 \le j \le N_x - 1,\ 0 \le m \le N_t - 1. \qquad (5.11)$$

Por lo tanto una vez que se calcula la solución en el nivel de tiempo $t = t_m$, la solución en el siguiente nivel del tiempo $t = t_{m+1}$ se puede encontrar de forma explícita. El esquema hacia

adelante (5.8)−(5.10) es un método explícito. Alternativamente, se puede reemplazar la derivada en el tiempo con una diferencia hacia atrás y seguir usando las diferencias centradas de segundo orden para la segunda derivada espacial. El esquema resultante es un esquema centrado en el espacio y tiempo hacia atrás:

$$\begin{aligned} \frac{v_j^m - v_j^{m-1}}{h_t} &= \nu \frac{v_{j+1}^m - 2v_j^m + v_{j-1}^m}{h_x^2} + f_j^m, \ 1 \leq j \leq N_x - 1, \, 1 \leq m \leq N_t, && (5.12) \\ v_0^m &= v_{N_x}^m = 0, \ 1 \leq m \leq N_t, && (5.13) \\ v_j^0 &= u_0 \left(x_j \right), \ 0 \leq j \leq N_x. && (5.14) \end{aligned}$$

La ecuación en diferencias (5.12) puede escribirse como

$$(1 + 2r) \, v_j^m - r \left(v_{j+1}^m + v_{j-1}^m \right) = v_j^{m-1} + h_t f_j^m, \ 1 \leq j \leq N_x - 1, \, 1 \leq m \leq N_t, \tag{5.15}$$

el cual se complementa con la condición de frontera de (5.13). Así, encontramos la solución en el nivel de tiempo $t = t_m$ a partir de la solución en $t = t_{m-1}$, tenemos que resolver un sistema lineal tridiagonal de orden $N_x - 1$. El esquema hacia atrás (5.12)−(5.14) es un método implícito. En los dos métodos anteriores, aproximamos la ecuación diferencial en $x = x_j$ y $t = t_m$. También podemos considerar la ecuación diferencial en $x = x_j$ y $t = t_{m-1/2}$, aproximando la derivada en el tiempo por diferencias centradas:

$$u_t \left(x_j, t_{m-1/2} \right) \approx \frac{u \left(x_j, t_m \right) - u \left(x_j, t_{m-1} \right)}{h_t}.$$

Además, la aproximación de la segunda derivada espacial por diferencias centradas de segundo orden:

$$u_{tt} \left(x_j, t_{m-1/2} \right) \approx \frac{u \left(x_{j+1}, t_{m-1/2} \right) - 2u \left(x_j, t_{m-1/2} \right) + u \left(x_{j-1}, t_{m-1/2} \right)}{h_t^2},$$

y luego aproximar los valores de un medio de tiempo, por los promedios

$$u \left(x_j, t_{m-1/2} \right) \approx \left[u \left(x_j, t_m \right) + u \left(x_j, t_{m-1} \right) \right] / 2,$$

etc. Como resultado llegamos al esquema de Crank-Nicolson:

$$\frac{v_j^m - v_j^{m-1}}{h_t} = \nu \frac{\left(v_{j+1}^m - 2v_j^m + v_{j-1}^m \right) + \left(v_{j+1}^{m-1} - 2v_j^{m-1} + v_{j-1}^{m-1} \right)}{2h_x^2} + f_j^{m-1/2},$$

$$\begin{aligned} & 1 \leq j \leq N_x - 1, \, 1 \leq m \leq N_t, && (5.16) \\ v_0^m &= v_{N_x}^m = 0, \ 1 \leq m \leq N_t, && (5.17) \\ v_j^0 &= u_0 \left(x_j \right), \ 0 \leq j \leq N_x. && (5.18) \end{aligned}$$

Aquí, $f_j^{m-1/2} = f \left(x_j, t_{m-1/2} \right)$, que puede ser sustituido por $\left(f_j^m + f_j^{m-1} \right) / 2$. La ecuación en

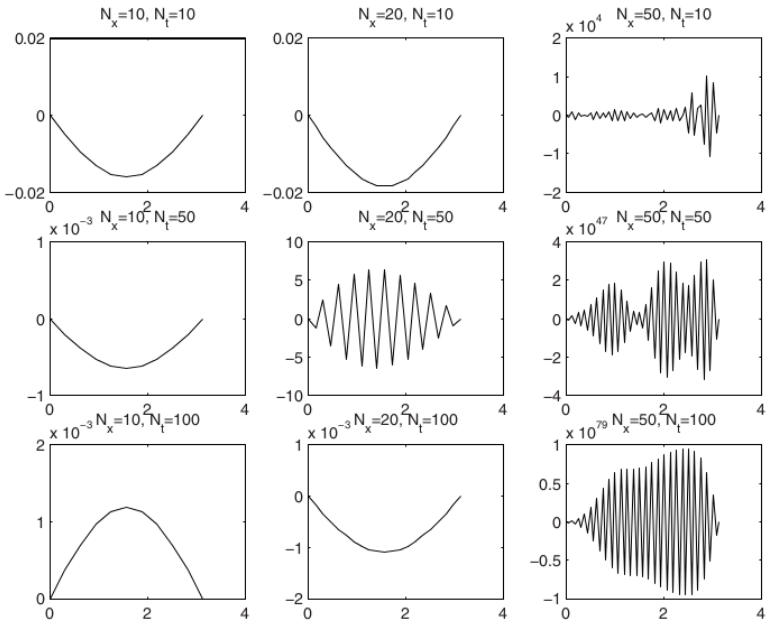
diferencias (5.16) se puede reescribir como

$$(1+r)\,v_j^m - \frac{r}{2}\left(v_{j+1}^m + v_{j-1}^m\right) = (1-r)\,v_j^{m-1} + \frac{r}{2}\left(v_{j+1}^{m-1} + v_{j-1}^{m-1}\right) + h_t f_j^{m-1/2}. \tag{5.19}$$

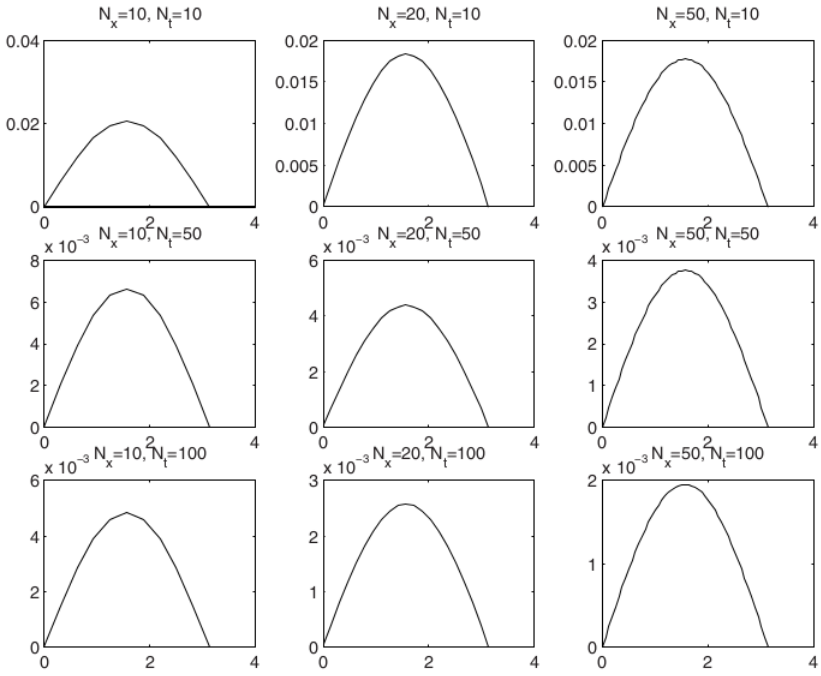
Vemos que el esquema de Crank -Nicolson es también un método implícito y para cada paso de tiempo necesitamos resolver un sistema lineal tridiagonal de orden $N_x - 1$.

Los tres esquemas derivados, parecen aproximaciones razonables para el problema de valor inicial- frontera $(5.5) - (5.7)$. Vamos a hacer algunos experimentos numéricos para ver si estos esquemas producen resultados útiles. Vamos a usar el esquema hacia adelante $(5.8) - (5.10)$ y el esquema de atrás $(5.12) - (5.14)$ para resolver el problema $(5.5) - (5.7)$ con $\nu = 1$, $f(x,t) = 0$ y $u_0(x) = \sin x$. Se puede comprobar que la solución exacta es $u(x,t) = \exp(-t)\sin(x)$. Consideramos los errores de la solución numérica en $t = 1$. Los resultados del esquema de Crank-Nicolson son cualitativamente similares al esquema de diferencias hacia atrás, pero las magnitudes de error son más pequeños, y por lo tanto se omiten .

En la siguiente figura se muestran los errores de la solución del esquema de avance correspondiente a varias combinaciones de los valores de N_x y N_t (o equivalentemente, h_x y h_t). La convergencia se observó sólo cuando N_x es sustancialmente menor que N_t (es decir, cuando h_t es sustancialmente menor que h_x).



La figura siguiente muestra los errores de soluciones del sistema hacia atrás correspondiente a los mismos valores de N_x y N_t. Observamos un buen patrón de convergencia. El error máximo de la solución disminuye a medida que N_x y N_t aumentan.



Naturalmente, un esquema de diferencias es útil sólo si el esquema es convergente, es decir, si puede proporcionar soluciones numéricas que se aproximan a la solución exacta. Un requisito necesario para la convergencia es la consistencia del sistema, es decir, el esquema de diferencias debe estar cerca de la ecuación diferencial en algún sentido. Sin embargo, la consistencia por sí sola no garantiza la convergencia, como lo vemos en los ejemplos numéricos anteriores. Desde el punto de vista de análisis teórico, en cada nivel de tiempo se produce un error, lo que representa la diferencia entre el esquema de diferencias y la ecuación diferencial. Desde el punto de vista de la aplicación informática, los valores numéricos y los cálculos numéricos están sujetas a errores de redondeo. Por lo tanto, es importante ser capaz de controlar la propagación de errores. La capacidad de controlar la propagación de errores se denomina estabilidad del esquema. Esperamos contar con la convergencia de los esquemas consistentes y estables. La teoría de Lax conocida por métodos de diferencias finitas va más allá de esto. La teoría dice que con las nociones definidas adecuadamente de consistencia, estabilidad y convergencia para un problema de la ecuación diferencial parcial bien definido, un esquema coherente es convergente si y sólo si es estable.

En la siguiente sección, se presenta una versión de la teoría de la equivalencia de Lax en la convergencia de los esquemas de diferencias.

5.2. Teorema de equivalencia de Lax

En esta sección, presentamos una versión del teorema de equivalencia de Lax para analizar los

métodos de diferencias en la solución de valor inicial o de problemas de valor de frontera. La teoría rigurosa se desarrolla en un marco abstracto. Para ayudar a entender la teoría, se utiliza el problema de la muestra $(5.5)-(5.7)$ con $f(x,t)=0$ para ilustrar la notación, suposiciones, definiciones y el resultado de equivalencia.

En primer lugar, presentamos un marco abstracto. Sea V un espacio de Banach, $V_0 \subset V$ un subespacio denso de V. Sea $L: V_0 \subset V \to V$ un operador lineal. El operador L es generalmente no acotado y puede ser pensado como un operador diferencial. Considere el problema de valor inicial

$$\begin{cases} \frac{du(t)}{dt} = Lu\left(t\right), \ \ 0 \leq t \leq T, \\ u\left(0\right) = u_0. \end{cases} \tag{5.20}$$

Este problema también representa un problema de valor de frontera con condiciones de contorno homogéneas cuando se incluyen en las definiciones del espacio V y el operador L. La siguiente definición da el significado de una solución del problema (5.20).

Definición 5.2 Una función $u:[0,T] \to V$ es una solución del problema de valor inicial (5.20) si para cualquier $t \in [0,T]$, $u\left(t\right) \in V_0$,

$$\lim_{\triangle t \to 0} \left\| \frac{1}{\triangle t}\left[u\left(t+\triangle t\right) - u\left(t\right)\right] - Lu\left(t\right) \right\| = 0, \tag{5.21}$$

y $u\left(0\right) = u_0$.

En la definición anterior, el límite en (5.21) se entiende que es el límite por la derecha en $t=0$ y el límite por la izquierda en $t=T$.

Definición 5.3 El problema de valor inicial (5.20) está bien planteado si para cualquier $u_0 \in V_0$, existe una única solución $u = u(t)$ y la solución depende de forma continua en el valor inicial: Existe una constante $c_0 > 0$ tal que si $u(t)$ y $\overline{u}(t)$ son las soluciones para los valores iniciales $u_0, \overline{u}_0 \in V_0$, entonces

$$\sup_{0 \leq t \leq T} \left\| u\left(t\right) - \overline{u}\left(t\right) \right\|_V \leq c_0 \left\| u_0 - \overline{u}_0 \right\|_V. \tag{5.22}$$

A partir de ahora, asumimos que el problema de valor inicial (5.20) está bien planteado. Se denota la solución como

$$u(t) = S(t)u_0, \quad u_0 \in V_0.$$

Usando la linealidad del operador L, se ve que el operador solución $S(t)$ es lineal. De la propiedad de dependencia continua (5.22), tenemos que

$$\begin{aligned} \sup_{0 \leq t \leq T} \left\| S\left(t\right)\left(u_0 - \overline{u}_0\right) \right\|_V &\leq c_0 \left\| u_0 - \overline{u}_0 \right\|_V, \\ \sup_{0 \leq t \leq T} \left\| S\left(t\right) u_0 \right\|_V &\leq c_0 \left\| u_0 \right\|_V, \quad \forall u_0 \in V_0. \end{aligned}$$

El operador $S(t) : V_0 \subset V \to V$ se puede extender de forma única a un operador lineal continuo $S(t) : V \to V$ con

$$\sup_{0\leq t\leq T} \|S(t)\|_V \leq c_0.$$

Definición 5.4 Para $u_0 \in V \setminus V_0$, llamamos $u(t) = S(t)u_0$ la solución generalizada del problema de valor inicial (5.20).

Ejemplo 5.5 Utilizamos el siguiente problema y sus aproximaciones en diferencias finitas para ilustrar el uso del marco abstracto de la sección:

$$\begin{cases} u_t = \nu u_{xx}, & \text{en } (0,\pi)\times(0,T), \\ u(0,t) = u(\pi,t) = 0, & 0 \leq t \leq T, \\ u(x,0) = u_0(t) & 0 \leq x \leq \pi. \end{cases} \tag{5.23}$$

Tomamos $V = C_0[0,\pi] = \{v \in C[0,\pi] \mid v(0) = v(\pi) = 0\}$, con la norma $\|.\|_{C[0,\pi]}$. Y elegimos

$$V_0 = \left\{ v \mid v(x) = \sum_{j=1}^{n} a_j \sin(jx),\ a_j \in \mathbb{R},\ n = 1,2,... \right\}. \tag{5.24}$$

Si $u_0 \in V_0$, entonces para cualquier entero positivo $n \geq 1$ y $b_1, ..., b_n \in \mathbb{R}$,

$$u_0(x) = \sum_{j=1}^{n} b_j \sin(jx). \tag{5.25}$$

Para este u_0, se verifica directamente que la solución es

$$u(x,t) = \sum_{j=1}^{n} b_j e^{-\nu j^2 t} \sin(jx). \tag{5.26}$$

Al utilizar el principio del máximo para la ecuación de calor,

$$\text{mín}\left\{0, \min_{0\leq x\leq\pi} u_0(x)\right\} \leq u(x,t) \leq \text{máx}\left\{0, \max_{0\leq x\leq\pi} u_0(x)\right\},$$

vemos que

$$\max_{0\leq x\leq\pi} |u(x,t)| \leq \max_{0\leq x\leq\pi} |u_0(x)| \quad \forall t \in [0,T].$$

Así, el operador solución $S(t) : V_0 \subset V \to V$ es acotado.

Entonces, para un $u_0 \in V$ general, el problema (5.23) tiene una solución única. Si $u_0 \in V$ tiene

una derivada continua a trozos en $[0,\pi]$, entonces,

$$u_0(x) = \sum_{j=1}^{\infty} b_j \sin(jx)$$

y la solución $u(t)$ puede ser expresada como

$$u(x,t) = S(t)\,u_0(x) = u_0(x) = \sum_{j=1}^{\infty} b_j e^{-\nu j^2 t} \sin(jx).$$

Regresamos al problema abstracto (5.20). Presentamos dos resultados, el primero es la continuidad en el tiempo de la solución generalizada y el segundo muestra que el operador solución $S(t)$ forma un semigrupo.

Proposición 5.6 Para cualquier $u_0 \in V$, la solución generalizada del problema de valor inicial (5.20) es continua en t.

Prueba. Sea $\{u_{0,n}\} \subset V_0$ una sucesión que converge a u_0 en V:

$$\|u_{0,n} - u_0\|_V \to 0 \text{ cuando } n \to \infty.$$

Sea $t_0 \in [0,T]$ fijo, y $t \in [0,T]$. Escribimos

$$\begin{aligned} u(t) - u(t_0) &= S(t)\,u_0 - S(t_0)\,u_0 \\ &= S(t)\,(u_0 - u_{0,n}) + [S(t) - S(t_0)]\,u_{0,n} - S(t_0)\,(u_0 - u_{0,n}). \end{aligned}$$

Entonces

$$\|u(t) - u(t_0)\|_V \le 2c_0 \|u_{0,n} - u_0\|_V + \|[S(t) - S(t_0)]\,u_{0,n}\|_V.$$

Dado cualquier $\varepsilon > 0$, elegimos n suficientemente grande tal que

$$2c_0 \|u_{0,n} - u_0\|_V < \frac{\varepsilon}{2}.$$

Para este n, usando (5.21) de la definición de la solución, tenemos un $\delta > 0$ de tal forma que

$$\|[S(t) - S(t_0)]\,u_{0,n}\|_V < \frac{\varepsilon}{2} \text{ para } |t - t_0| < \delta.$$

Entonces para $t \in [0,T]$ con $|t - t_0| < \delta$, tenemos que $\|u(t) - u(t_0)\|_V \le \varepsilon$.

♢

Proposición 5.7 Suponga que el problema (5.20) está bien planteado. Entonces, para todo $t_1, t_0 \in [0,T]$ tal que $t_1 + t_0 \le T$, tenemos $S(t_1 + t_0) = S(t_1)S(t_0)$.

Prueba. La solución del problema (5.20) es $u(t) = S(t)u_0$. Tenemos que $u(t_0) = S(t_0)u_0$ y

$S(t)u(t_0)$ es la solución de la ecuación diferencial en $[t_0, T]$ con la condición inicial $u(t_0)$ en t_0. Por la unicidad de la solución,

$$S(t)u(t_0) = u(t + t_0),$$

es decir,

$$S(t_1)S(t_0)u_0 = S(t_1 + t_0)u_0.$$

Como $u_0 \in V$ es arbitraria, $S(t_1 + t_0) = S(t_1)S(t_0)$.

$\diamondsuit$

Ahora introducimos un método de diferencias finitas definido por un parámetro de la familia de los operadores lineales uniformemente acotados

$$C(\triangle t) : V \to V, \quad 0 < \triangle t < \triangle_0.$$

Aquí $\triangle_0 > 0$ es un número fijo. La familia $\{C(\triangle t)\}_{0<\triangle t<\triangle_0}$ se dice que es uniformemente acotada si existe una constante c tal que

$$\|C(\triangle t)\| \leq c \quad \forall \triangle t \in (0, \triangle_0].$$

La solución aproximada se define entonces por

$$u_{\triangle t}(m\triangle t) = C(\triangle t)^m u_0, \quad m = 1, 2, ...$$

Definición 5.8 (Consistencia) El método en diferencias es consistente si existe un subespacio denso V_c de V tal que para todo $u_0 \in V_c$, para la correspondiente solución u del problema de valor inicial (5.20), tenemos

$$\lim_{\triangle t \to 0} \left\| \frac{1}{\triangle t} [C(\triangle t)\, u(t) - u(t + \triangle t)] \right\| = 0 \quad \text{uniformemente en } [0, T].$$

Ejemplo 5.9 (continuación del Ejemplo 5.5) Consideremos ahora el método hacia adelante y el método hacia atrás del Ejemplo 5.1 para el mismos problema (5.32). Para el método hacia adelante, definimos el operador $C(\triangle t)$ por la fórmula

$$C(\triangle t)\, v(x) = (1 - 2r)\, v(x) + r\, [v(x + \triangle x) + v(x - \triangle x)],$$

donde $\triangle x = \sqrt{\nu \triangle t / r}$ y si $x \pm \triangle x \notin [0, \pi]$, entonces la función v se extiende con singularidad con periodo 2π. Identificamos $\triangle t$ con h_t y $\triangle x$ con h_x. Entonces $C(\triangle t) : V \to V$ es un operador lineal y cumple que

$$\|C(\triangle t)\, v\|_V \leq (|1 - 2r| + 2r) \|v\|_V \quad \forall\, v \in V.$$

Así

$$\|C(\triangle t)\|_V \leq |1-2r| + 2r, \tag{5.27}$$

y la familia $\{C(\triangle t)\}$ es uniformemente acotada. El método en diferencias es

$$u_{\triangle t}(t_m) = C(\triangle t)\, u_{\triangle t}(t_{m-1}) = C(\triangle t)^m u_0$$

o

$$u_{\triangle t}(\cdot, t_m) = C(\triangle t)^m u_0(\cdot).$$

Observe que en esta forma, el método de diferencias genera una solución aproximada $u_{\triangle t}(x,t)$ que está definida para $x \in [0,\pi]$ y $t = t_m$, $m = 0,1,...,N_t$. Dado que

$$\begin{aligned}
u_{\triangle t}(x_j, t_{m+1}) &= (1-2r)\, u_{\triangle t}(x_j, t_m) + r\left[u_{\triangle t}(x_{j-1}, t_m) + u_{\triangle t}(x_{j+1}, t_m)\right], \\
&\qquad 1 \leq j \leq N_x - 1,\ 0 \leq m \leq N_t - 1, \\
u_{\triangle t}(0, t_m) &= u_{\triangle t}(N_x, t_m) = 0,\ \ 0 \leq m \leq N_t, \\
u_{\triangle t}(t_j, 0) &= u_0(x_j),\ \ 0 \leq j \leq N_x,
\end{aligned}$$

vemos que la relación entre la solución aproximada $u_{\triangle t}$ y la solución v definida por el esquema de diferencias ordinario $(5.8) - (5.10)$ (con $f_j^m = 0$) es

$$u_{\triangle t}(x_j, t_m) = v_j^m. \tag{5.28}$$

En cuanto a la consistencia, tomamos $V_c = V_0$. Para la función de valor inicial (5.25), tenemos la fórmula (5.26) para la solución la cual es infinitamente suave. Ahora, utilizando la expansión de Taylor en (x,t), tenemos

$$\begin{aligned}
C(\triangle t)\, u(x,t) - u(x, t+\triangle t) &= (1-2r)\, u(x,t) + r\left[u(x+\triangle x, t) + u(x-\triangle x, t)\right] - u(x, t+\triangle t) \\
&= (1-2r)\, u(x,t) + r\left[2u(x,t) + u_{xx}(x,t)(\triangle x)^2\right] \\
&\quad + \frac{r}{4!}\left[u_{xxxx}(x+\theta_1\triangle x, t) + u_{xxxx}(x-\theta_2\triangle x, t)\right](\triangle x)^4 \\
&\quad - u(x,t) - u_t(x,t)\triangle t - \frac{1}{2}u_{tt}(x, t+\theta_3\triangle t)(\triangle t)^2 \\
&= -\frac{1}{2}u_{tt}(x, t+\theta_3\triangle t)(\triangle t)^2 \\
&\quad - \frac{\nu^2}{24r}\left[u_{xxxx}(x+\theta_1\triangle x, t) + u_{xxxx}(x-\theta_2\triangle x, t)\right](\triangle t)^2,
\end{aligned}$$

donde $\theta_1, \theta_2, \theta_3 \in (0,1)$. Entonces

$$\left\| \frac{1}{\triangle t}\left[C(\triangle t)\, u(t) - u(t+\triangle t)\right] \right\| \leq c\triangle t$$

y tenemos la consistencia del esquema.

Volviendo al caso general.

Definición 5.10 (Convergencia) El método de diferencias finitas es convergente si para cualquier t fijo en $[0,T]$ y cualquier $u_0 \in V$, tenemos

$$\lim_{\triangle t_i \to 0} \left\| \left[C\left(\triangle t_i\right)^{m_i} - S\left(t\right) \right] u_0 \right\| = 0,$$

donde $\{m_i\}$ es una sucesión de enteros y $\{\triangle t_i\}$ es una sucesión de tamaños de paso tal que $\lim_{i\to\infty} m_i \triangle t_i = t$.

Definición 5.11 (Estabilidad) El método de diferencias finitas es estable si los operadores

$$\left\{ C\left(\triangle t\right)^m \mid 0 < \triangle t < \triangle_0,\, m\triangle t \leq T \right\}$$

son uniformemente acotados, es decir, existe una constante $M_0 > 0$ tal que

$$\left\| C\left(\triangle t\right)^m \right\|_{V\to V} \leq M_0 \;\; \forall m:\; m\triangle t \leq T,\, \forall \triangle t \leq \triangle_0.$$

Vemos ahora el resultado central.

Teorema 5.12 (Teorema de Equivalencia de Lax) Supongamos que el problema de valor inicial (5.20) está bien planteado. Entonces, para un método de diferencias consistente, la estabilidad es equivalente a la convergencia.

Prueba. ($\Rightarrow$) Consideremos el error

$$C\left(\triangle t\right)^m u_0 - u\left(t\right) = \sum_{j=1}^{m-1} C\left(\triangle t\right)^j \left[C\left(\triangle t\right) u\left(\left(m-1-j\right)\triangle t\right) - u\left(\left(m-j\right)\triangle t\right) \right] + u\left(m\triangle t\right) - u\left(t\right).$$

Primero asumimos que $u_0 \in V_c$. Entonces como el método es estable,

$$\left\| C\left(\triangle t\right)^m u_0 - u\left(t\right) \right\| \leq M_0 m\triangle t \sup_t \left\| \frac{C\left(\triangle t\right) u\left(t\right) - u\left(t+\triangle t\right)}{\triangle t} \right\| + \left\| u\left(m\triangle t\right) - u\left(t\right) \right\|. \quad (5.29)$$

Por la continuidad, $\left\| u\left(m\triangle t\right) - u\left(t\right) \right\| \to 0$, y por la consistencia,

$$\sup_t \left\| \frac{C\left(\triangle t\right) u\left(t\right) - u\left(t+\triangle t\right)}{\triangle t} \right\| \to 0,$$

por tanto tenemos la convergencia por (5.29).

Consideremos ahora la convergencia para el caso general donde $u_0 \in V$. Tenemos la secuencia $\{u_{0,n}\} \subset V_0$ tal que $u_{0,n} \to u_0$ en V. Escribimos

$$C\left(\triangle t\right)^m u_0 - u\left(t\right) = C\left(\triangle t\right)^m \left(u_0 - u_{0,n}\right) + \left[C\left(\triangle t\right)^m - S\left(t\right) \right] u_{0,n} - S\left(t\right)\left(u_0 - u_{0,n}\right),$$

y obtenemos

$$\left\| C\left(\triangle t\right)^m u_0 - u\left(t\right) \right\| \leq \left\| C\left(\triangle t\right)^m \left(u_0 - u_{0,n}\right) \right\| + \left\| \left[C\left(\triangle t\right)^m - S\left(t\right) \right] u_{0,n} \right\| + \left\| S\left(t\right)\left(u_0 - u_{0,n}\right) \right\|.$$

Dado que el problema de valor inicial (5.20) está bien planteado, y el método es estable,

$$\left\| C\left(\triangle t\right)^m u_0 - u\left(t\right) \right\| \leq c \left\| u_0 - u_{0,n} \right\| + \left\| \left[C\left(\triangle t\right)^m - S\left(t\right) \right] u_{0,n} \right\|.$$

Dado cualquier $\varepsilon > 0$, existe un n suficientemente grande de tal manera que

$$c \left\| u_0 - u_{0,n} \right\| < \frac{\varepsilon}{2}$$

Para este n, sea $\triangle t$ suficientemente pequeño,

$$\|[C(\triangle t)^m - S(t)]\, u_{0,n}\| < \frac{\varepsilon}{2} \ \forall \triangle t \text{ pequeño}, \ |m\triangle t - t| < \triangle t.$$

Entonces obtenemos la convergencia.

($\Leftarrow$) Supongamos que el método no es estable. Luego existen sucesiones $\{\triangle t_k\}$ y $\{m_k\}$ tal que $m_k \triangle t_k \leq T$ y

$$\lim_{k\to\infty} \|C(\triangle t_k)^{m_k}\| = \infty.$$

Como $\triangle t_k \leq \triangle_0$, podemos asumir que la sucesión $\{\triangle t_k\}$ es convergente. Si la sucesión $\{m_k\}$ es acotada, entonces

$$\sup_k \|C(\triangle t_k)^{m_k}\| \leq \sup_k \|C(\triangle t_k)\|^{m_k} \leq \infty.$$

Esto es una contradicción. Así $m_k \to \infty$ y $\triangle t_k \to 0$ cuando $k \to \infty$.

Por la convergencia del método,

$$\sup_k \|C(\triangle t_k)^{m_k} u_0\| \leq \infty \ \ \forall u_0 \in V.$$

Aplicando el teorema que dice que dada una secuencia $\{L_n\}$ de operadores lineales de un espacio de Banach V a un espacio normado W, asumiendo que para todo $v \in V$ y que la secuencia $\{L_n v\}$ es acotada entonces $\sup_n \|L_n\| < \infty$, tenemos

$$\lim_{k\to\infty} \|C(\triangle t_k)^{m_k}\| < \infty,$$

contradiciendo la suposición que el método no es estable.

$\diamondsuit$

Corolario 5.13 (Orden de convergencia) Bajo los supuestos del Teorema 5.12, si u es una solución con valor inicial $u_0 \in V_c$, que satisface

$$\sup_{0\leq t\leq T} \left\| \frac{C(\triangle t)\, u(t) - u(t+\triangle t)}{\triangle t} \right\| \leq c(\triangle t)^k \ \ \forall \triangle t \in (0, \triangle_0],$$

entonces tenemos la estimación del error

$$\|C(\triangle t)^m u_0 - u(t)\| \leq c(\triangle t)^k,$$

donde m es un entero positivo con $m\triangle t = t$.

Prueba. El error estimado se sigue inmediatamente de (5.29).

$\diamondsuit$

Ejemplo 5.14 (continuación del ejemplo 5.9) Apliquemos el Teorema de equivalencia de Lax para el esquema hacia adelante, suponemos $r \leq 1/2$. Entonces de acuerdo con (5.27), $\|C(\triangle t)\| \leq 1$ y así

$$\|C(\triangle t)^m\| \leq 1, \ \ m = 1, 2, ...$$

Así, bajo la condición $r \leq 1/2$, el esquema hacia delante es estable. Como el esquema es consistente, tenemos la convergencia

$$\lim_{\triangle t_i \to 0} \left\| u_{\triangle t}\left(\cdot, m_i \triangle t_i\right) - u\left(\cdot, t\right) \right\|_V = 0, \tag{5.30}$$

donde $\lim_{\triangle t_i \to 0} m_i \triangle t_i = t$.

En realidad, $\|C\left(\triangle t\right)\| = |1 - 2r| + 2r$ y $r \leq 1/2$ es una condición necesaria y suficiente para la estabilidad y luego para la convergencia.

Por la relación (5.28), para la solución de diferencias finitas $\left\{v_j^m\right\}$ se define en $(5.8) - (5.10)$ con $f_j^m = 0$, tenemos la convergencia

$$\lim_{h_t \to 0} \max_{0 \leq j \leq N_x} \left| v_j^m - u\left(x_j, t\right) \right| = 0,$$

donde m depende de h_t y $\lim_{h_t \to 0} m h_t = t$.

Como necesitamos una condición ($r \leq 1/2$ en este caso) para la convergencia, el esquema hacia delante se dice que es condicionalmente estable y condicionalmente convergente.

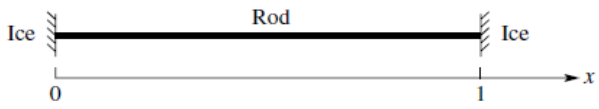
6. Capítulo 3. Problemas parabólicos

6.1. Problema modelo de la ecuación de calor

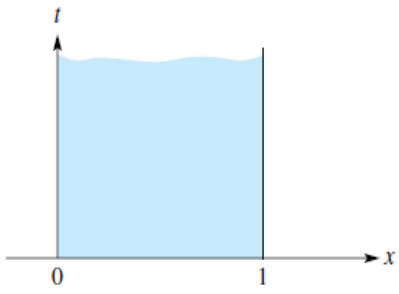
Consideramos un problema modelo, para introducir algunas de las ideas esenciales. En él tenemos la ecuación de calor en una variable acompañada de las condiciones de frontera adecuadas para un determinado fenómeno físico:

$$\begin{cases} \frac{\partial^2}{\partial x^2} u(x,t) &= \frac{\partial}{\partial t} u(x,t), \\ u(0,t) &= u(1,t) = 0, \\ u(x,0) &= \sin \pi x. \end{cases} \tag{6.1}$$

Esta ecuación gobierna la temperatura $u(x,t)$ en una varilla delgada de longitud 1 cuando los extremos están a una temperatura 0, bajo el supuesto que la temperatura inicial de la varilla está dada por la función $\sin \pi x$, como podemos ver en la figura:



En el plano xt, la región en la cual se busca la solución esta descrita por las desigualdades $0 \leq x \leq 1$ y $t \geq 0$. En la frontera de esta región, los valores de u son conocidos:



6.2. Método de diferencias finitas.

Un enfoque fundamental en la solución de tal problema es el método de diferencias finitas. Se procede mediante el reemplazo de las derivadas en la ecuación por diferencias finitas. Dos fórmulas son útiles en este contexto:

$$f'(x) \approx \frac{1}{h}\left[f(x+h) - f(x)\right],$$

$$f''(x) \approx \frac{1}{h^2}\left[f(x+h) - 2f(x) + f(x-h)\right].$$

Si aplicamos estas fórmulas a la ecuación diferencial (6.1), con posiblemente diferentes pasos de longitud h y k, el resultado es

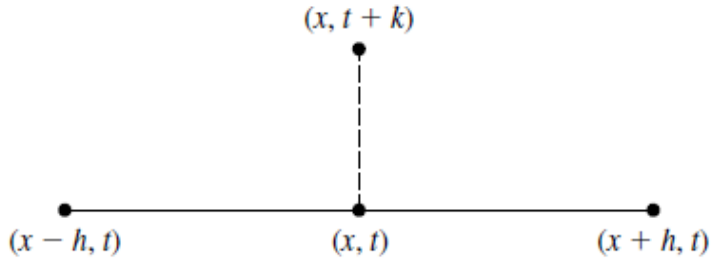
$$\frac{1}{h^2}\left[u(x+h,t) - 2u(x,t) + u(x-h,t)\right] = \frac{1}{k}\left[u(x,t+k) - u(x,t)\right] \tag{6.2}$$

esta ecuación es interpretada ahora como un medio para avanzar a la solución paso a paso en la variable t. Es decir, si $u(x,t)$ es conocida en $0 \leq x \leq 1$ y $0 \leq t \leq t_0$, la ecuación (6.2) nos permite evaluar la solución para $t = t_0 + k$.

La ecuación (6.2) puede ser reescrita de la forma

$$u(x,t+k) = \sigma u(x+h,t) + (1-2\sigma)u(x,t) + \sigma u(x-h,t), \text{ donde } \sigma = \frac{k}{h^2} \tag{6.3}$$

Un croquis que muestra la ubicación de los cuatro puntos que participan en esta ecuación se da en la figura:



Como la solución se conoce en la frontera de la región, es posible calcular una solución aproximada en el interior de la región usando sistemáticamente la ecuación (6.3) porque la ecuación (6.2) es solo una diferencia finita análoga a la ecuación (6.1).

Para obtener una solución aproximada en la computadora, nosotros seleccionamos valores para h y k y usamos la ecuación (6.3). Usando este algoritmo, podemos continuar con la solución indefinidamente en la variable t, los cálculos involucran solo los valores anteriores de t.

6.3. Pseudocódigo para el método explícito

Para mayor simplicidad, seleccionamos $h = 0.1$ y $k = 0.005$, así el coeficiente $\sigma = 0.5$, esta elección hace que el coeficiente $1 - 2\sigma$ sea cero. Nuestro pseudocódigo primero toma $u(ih, 0)$ para $0 \leq i \leq 10$ por que se conocen los valores en la frontera, luego se calcula $u(ih, k)$ para $0 \leq i \leq 10$ usando la ecuación (6.3) y los valores de frontera $u = (0, t) = u(1, t) = 0$. Este proceso se continua hasta que t alcanza el valor 0.1. Las matrices u y v se usan para almacenar la solución aproximada en t y $t + h$, respectivamente. Como la solución analítica al problema es $u(x, t) = \exp(-\pi^2 t) \sin(\pi x)$, el error puede se impreso en cada paso.

El procedimiento descrito es un ejemplo de un método explícito. Los valores aproximados de $u(x, t + k)$ son calculados explícitamente en términos de $u(x, t)$.

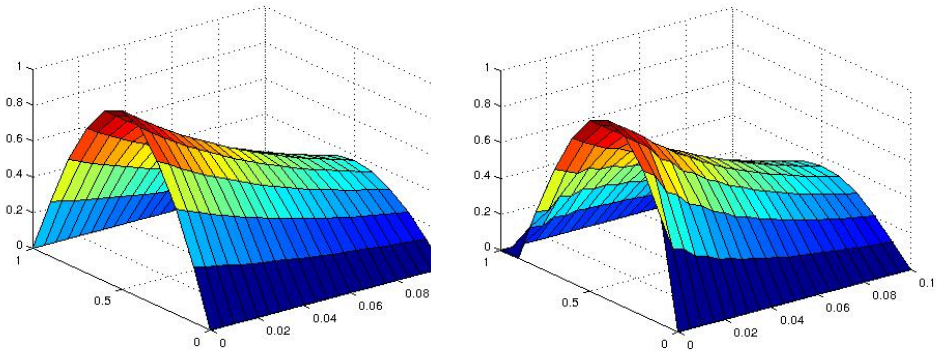
Como h debe ser pequeño para representar de manera precisa la derivada por la fórmula de diferencias finitas, el correspondiente valor de k debe ser pequeño también. Valores tales como $h = 0.1$ y $k = 0.005$ son representativos, así como $h = 0.01$ y $k = 0.0005$. Con valores pequeños de k, se necesitan muchos cálculos para avanzar bastante en la variable t.

A continuación se presenta el código para su implementación en Octave/MATLAB:

```
%solucion aproximada de la EDP parabolica
n=10; m=20; h=0.1; k=0.005;

u=zeros(n+1,m+1); v=zeros(n+1,m+1); er=zeros(n+1,m+1);
u(2:n-1,1)=sin(pi*(2:n-1)*h);
v(2:n-1,1)=sin(pi*(2:n-1)*h);

t=0;
for j=2:m+1;
    for i=2:n;
        u(i,j)=(k/h^(2))*(u(i-1,j-1)+u(i+1,j-1));
        v(i,j)= exp(-pi^2*t)*sin(pi*i*h);
        er(i,j)=v(i,j)-u(i,j);
    end
    t=j*k;
end
disp(u)
disp(v)
disp(er)

x=0:h:1; y=0:k:k*20;
[a b]=meshgrid(y,x);
surf(a, b, u)
```

```
%grafica de la sol. exacta de la EDP parabolica
a=0:0.1:1; b=0:.005:.005*20;
[x y]=meshgrid(a,b);
z=exp(-pi^2.*y).*sin(pi.*x);
%intercambie el orden de x e y
surf(y,x,z)
```

El cual devuelve la aproximación, el valor exacto y el error cada valor en distintas matrices, se grafican las aproximación y la exacta en la misma figura:



Si se juntan ambos gráficos, estos claramente coinciden en sus características generales se alcanza una temperatura máxima en el centro de la barra y ésta se escapa por los bordes donde la temperatura es de cero grados.

6.4. Método de Crank-Nicolson

Un procedimiento alternativo del tipo método implícito se conoce con el nombre de sus inventores, John Crank y Phillis Nicolson, y se basa en una simple variante de la ecuación original (6.1)

$$\frac{1}{h^2}\left[u\left(x+h,t\right)-2u\left(x,t\right)+u\left(x-h,t\right)\right]=\frac{1}{k}\left[u\left(x,t\right)-u\left(x,t-k\right)\right] \tag{6.4}$$

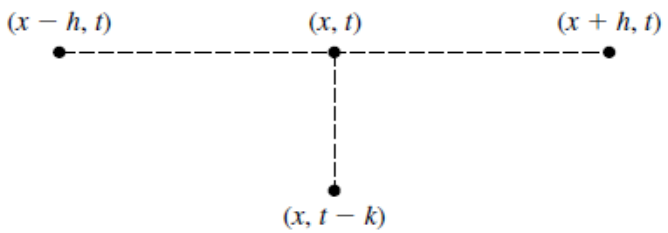
si la solución numérica de la malla de puntos $x = ih$, $t = jk$ se ha obtenido hasta cierto nivel en la variable t, la ecuación anterior gobierna los valores de u en el siguiente nivel de t.

Además la ecuación anterior se puede escribir como

$$-u\left(x-h,t\right)+ru\left(x,t\right)-u\left(x+h,t\right)=s\,u\left(x,t-k\right) \tag{6.5}$$

donde $r = 2 + s$ y $s = \frac{h^2}{k}$.

La localización de los cuatro puntos en esta ecuación se muestra en la figura:



En el nivel t, u es desconocido, pero en el nivel $(t - k)$, u es conocido, entonces podemos intro-

ducir incógnitas $u_i = u(ih, t)$ y cantidades conocidas $b_i = su(ih, t-k)$ y escribir la ecuación (6.5) en forma matricial:

$$\begin{bmatrix} r & -1 & & & & \\ -1 & r & -1 & & & \\ & -1 & r & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & r & -1 \\ & & & & -1 & r \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-2} \\ u_{n-1} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-2} \\ b_{n-1} \end{bmatrix} \tag{6.6}$$

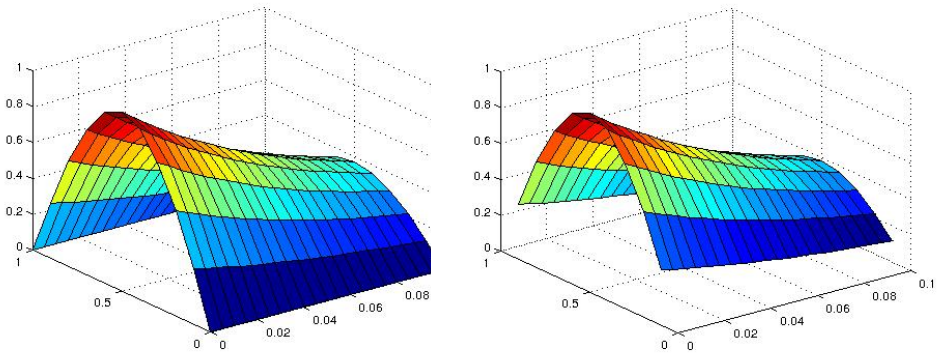
se ha utilizado el supuesto que $u(0,t) = u(1,t) = 0$. También $h = 1/n$. El sistema de ecuaciones es tridiagonal y diagonal dominante por que $|r| = 2 + h^2/k > 2$.

Pseudocódigo para el Método de Crank-Nicolson

A continuación llevamos a cabo el método de Crank-Nicolson para el problema modelo del inicio, escrito en Octave/MATLAB:

```
n=10; m=20; h=0.1; k=0.005;
s=h^2/k; r=2+s;
c=zeros(1,n-1); d=zeros(1,n-1); u=zeros(n-1,m); v=zeros(1,n-1);
u(:,1)=sin(pi*h*(1:n-1)');
for j=1:m
    v=s*u(:,j);
    u(:,j+1)=tri(-ones(1,n-1),r*ones(1,n-1),-ones(1,n-1),v);
    t=j*k;
end
disp(u)
x=0.1:h:0.9; y=0:k:k*20;
[a b]=meshgrid(y,x);
surf(a, b, u)
```

El cual devuelve sólo la aproximación y graficando junto con la solución exacta tenemos:



En él, $h = 0.1$, $k = h^2/2$, y la solución se continúa hasta $t = 0.1$. Los valores $r = 4$ y $s = 2$. Calculamos solo los valores de u para puntos interiores de cada linea horizontal. Para los puntos de la frontera tenemos $u(0,t) = u(1,t) = 0$.

Se usaron los mismos valores para k y h en el pseudocódigo de los dos métodos (explícito y Crank-Nicolson), por lo que puede hacerse un comparación en las salidas.

6.5. Versión alternativa del método de Crank

Otra versión del método de Crank Nicolson se obtiene de la siguiente manera:
Las diferencias centrales de $\left(x, t - \frac{1}{2}k\right)$ en la ecuación

$$\frac{1}{h^2}\left[u\left(x+h,t\right) - 2u\left(x,t\right) + u\left(x-h,t\right)\right] = \frac{1}{k}\left[u\left(x,t\right) - u\left(x,t-k\right)\right]$$

producen

$$\frac{1}{h^2}\left[u\left(x+h,t-\frac{1}{2}k\right) - 2u\left(x,t-\frac{1}{2}k\right) + u\left(x-h,t-\frac{1}{2}k\right)\right] = \frac{1}{k}\left[u\left(x,t\right) - u\left(x,t-k\right)\right]$$

como los valores de u son conocidos solo para los enteros múltiplos de k, los términos tales como $u\left(x, t - \frac{1}{2}k\right)$ son reemplazados por el promedio de los valores de u en los puntos adyacentes de la malla, esto es;

$$u\left(x,t-\frac{1}{2}k\right) \approx \frac{1}{2}\left[u\left(x,t\right) + u\left(x,t-k\right)\right]$$

entonces nosotros tenemos

$$\frac{1}{2h^2}\left[u\left(x+h,t\right) - 2u\left(x,t\right) + u\left(x-h,t\right) + u\left(x+h,t-k\right) - 2u\left(x,t-k\right) + u\left(x-h,t-k\right)\right] = \frac{1}{k}\left[u\left(x,t\right) - u\left(x,t-k\right)\right],$$

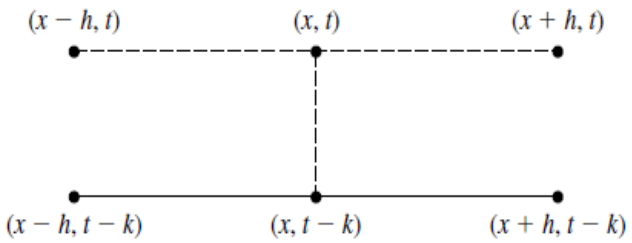
la forma computacional de esta ecuación es

$$-u\left(x-h,t\right) + 2\left(1+s\right)u\left(x,t\right) - u\left(x+h,t\right) = u\left(x-h,t-k\right) + 2\left(s-1\right)u\left(x,t-k\right) + u\left(x+h,t-k\right) \tag{6.7}$$

donde

$$s = \frac{h^2}{s} \equiv \frac{1}{\sigma}$$

Los seis puntos en esta ecuación se muestran en la figura:



Esto conduce al sistema tridiagonal de la forma (6.6) con $r = 2(1+s)$ y

$$b_i = u((i-1)h, t-k) + 2(s-1)u(ih, t-k) + u((i+1)h, t+k).$$

6.6. Estabilidad

La esencia del método explicito es la ecuación (6.3)

$$u(x, t+k) = \sigma u(x+h, t) + (1-2\sigma)u(x, t) + \sigma u(x-h, t)$$

la cual muestra como los valores de u para $t+k$ dependen de los valores de u en el paso del tiempo anterior, t. Si introducimos los valores de u en la malla escribiendo $u_{ij} = u(ih, jk)$, entonces podemos reunir todos los valores para un t-nivel en un vector $v^{(j)}$ como sigue:

$$v^{(j)} = [u_{0j}, u_{1j}, u_{2j}, ..., u_{nj}]^T,$$

y nuestra ecuación inicial ahora podría escribirse de la forma $u_{i,j+1} = \sigma u_{i+1,j} + (1-2\sigma)u_{ij} + \sigma u_{i-1,j}$.

Esta ecuación muestra como $v^{(j+1)}$ es obtenido de $v^{(j)}$. Esto es simplemente $v^{(j+1)} = Av^{(j)}$, donde A es la matriz cuyos elementos son

$$\begin{bmatrix} 1-2\sigma & \sigma & & & & & \\ \sigma & 1-2\sigma & \sigma & & & & \\ & \sigma & 1-2\sigma & \sigma & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \sigma & 1-2\sigma & \sigma & \\ & & & & \sigma & 1-2\sigma & \end{bmatrix}$$

Las ecuaciones nos dicen que $v^{(j)} = Av^{(j-1)} = A^2v^{(j-2)} = A^3v^{(j-3)} = ... = A^jv^0$.

A partir de las consideraciones físicas, la temperatura en la barra debería aproximarse a cero. Después de todo, el calor se pierde a través de los extremos de la varilla, que se mantuvieron a temperatura cero. Por lo tanto $A^j v^0$ debería converger a cero cuando $j \to \infty$.

7. Capítulo 4. Problemas hiperbólicos

7.1. La ecuación de onda 1D

El prototipo de una ecuación diferencial parcial hiperbólica lineal es la ecuación unidimensional

$$u_t + au_x = 0, \tag{7.1}$$

donde a es una contante, t representa el tiempo y x representa la variable espacial. Para que esta ecuación sea resoluble, las condiciones iniciales de la misma deben ser establecidas. En este caso es dado $u(0,x)$ es igual a una función $u_0(x)$. La incógnita es $u(t,x)$ para todo valor de t positivo.

La solución a la ecuación es $u(t,x) = u_0(x-at)$. Observe que esta solución puede fácilmente ser verificada reemplazándola en la ecuación $u_t + au_x = 0$ (esto prueba la existencia de una solución a la ecuación $u_t + au_x = 0$).

Analizando la solución podemos observar que:

- La solución en cualquier tiempo $t = T$ es una copia desplazada de la función original $u_0(x)$.
- El desplazamiento será hacia la derecha si a es positiva y hacia la izquierda si a es negativa.
- El tamaño de este desplazamiento es $|a|T$, esto significa que la solución en el punto (t,x) depende exclusivamente del valor de $\xi = x - at$.
- Las lineas en el plano (t,x) sobre la cual $x-at$ es constante es denominada de característica, y además, sobre la linea característica la función $u(t,x)$ tienen un valor constante igual a $u_0(\xi)$.
- Note que $a = (x-\xi)/t$ tiene dimensión de distancia sobre tiempo. Así podemos asociar a una velocidad: la velocidad de propagación. La ecuación modela una onda que se propaga con una velocidad a sin cambio en su forma.

Un hecho importante en la solución de ecuaciones diferenciales hiperbólicas es que la ecuación $u_t + au_x = 0$ solo tiene sentido si $u(t,x)$ es diferenciable, sin embargo la solución presentada $u_0(x-at)$ depende de este requerimiento.

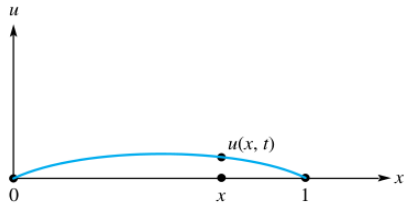
7.2. Problema modelo de la ecuación de onda

La ecuación de onda con una variable en el espacio es

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \tag{7.2}$$

la cual gobierna la vibración de una cuerda (vibración transversal en un plano) o la vibración en una varilla (vibración longitudinal). Es un ejemplo de una ecuación diferencial lineal de segundo orden del tipo hiperbólico. Si se utiliza la ecuación (7.2) para modelar la vibración de una cuerda, entonces $u(x,t)$ representa la desviación en el tiempo t de un punto de la cuerda cuya coordenada x parte del reposo.

Para plantear un problema modelo, suponemos que los puntos de la cuerda tienen coordenadas x en el intervalo $0 \leq x \leq 1$.

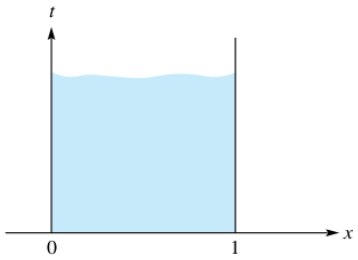


Luego supongamos que en el instante $t = 0$, las deflexiones satisfacen las ecuaciones $u(x,0) = f(x)$ y $u_t(x,0) = 0$. Supongamos también que los extremos de la cuerda se mantienen fijos. Entonces $u(0,t) = u(1,t) = 0$. Un problema de valores de frontera completamente definido, entonces, es

$$\begin{cases} u_{tt} - u_{xx} & = 0 \\ u\,(x,0) & = f\,(x) \\ u_t\,(x,0) & = 0 \\ u\,(0,t) & = u\,(1,t) = 0 \end{cases} \tag{7.3}$$

La región en el plano xt, donde se busca una solución es la franja semi-infinita definida por las desigualdades $0 \leq x \leq 1$ y $t \geq 0$.

Como en el problema de conducción de calor de, los valores de la función desconocida se prescriben en el límite de la región mostrada



7.2.1. Solución analítica

El problema modelo (7.3) es tan simple que puede ser resuelto de inmediato. En efecto, la solución es

$$u(x,t) = \frac{1}{2}\left[f(x+t) + f(x-t)\right] \tag{7.4}$$

A condición de que una función f posee dos derivadas y se ha extendido a toda la línea real, por definición se tiene$f(-x) = -f(x)$ y $f(x+2) = f(x)$.

Para comprobar que la ecuación (7.4) es una solución, calculamos las derivadas usando la regla de la cadena:

$$u_x = \tfrac{1}{2}\left[f'(x+t) + f'(x-t)\right] \qquad u_t = \tfrac{1}{2}\left[f'(x+t) - f'(x-t)\right],$$
$$u_{xx} = \tfrac{1}{2}\left[f''(x+t) + f''(x-t)\right] \qquad u_{tt} = \tfrac{1}{2}\left[f''(x+t) + f''(x-t)\right].$$

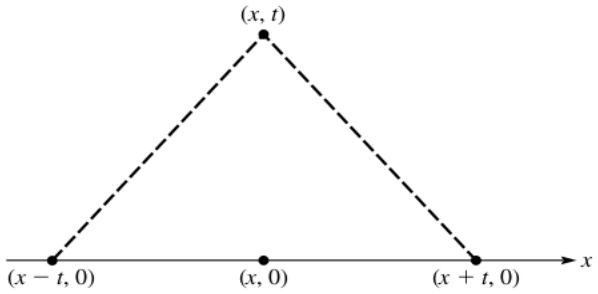
Obviamente $u_{tt} = u_{xx}$. También $u(x,0) = f(x)$. Por tanto, tenemos $u_t(x,0) = \frac{1}{2}\left[f'(x) - f'(x)\right] = 0$.

Al consultar las condiciones en los extremos, utilizamos las fórmulas por el cual se extendió f:

$$\begin{aligned} u(0,t) &= \frac{1}{2}\left[f(t) + f(-t)\right] = 0 \\ u(1,t) &= \frac{1}{2}\left[f(1+t) + f(1-t)\right] \\ &= \frac{1}{2}\left[f(1+t) - f(1-t)\right] \\ &= \frac{1}{2}\left[f(1+t) - f(t-1+2)\right] = 0. \end{aligned}$$

La extensión de f en su dominio original a toda la recta real hace que sea una función periódica impar de periodo 2. Impar significa que $f(-x) = -f(x)$ y la periodicidad es expresada por $f(x+2) = f(x)$, para todo x.

Para calcular $u(x,t)$, necesitamos conocer f y sólo dos puntos del eje x, que son $x+t$ y $x-t$ como en la figura siguiente



7.2.2. Solución numérica

El problema modelo se utiliza aquí para ilustrar de nuevo el principio de la solución numérica. La elección de tamaños de paso h y k para x y t, respectivamente, y el uso de las aproximaciones conocidas para las derivadas, tenemos de la ecuación (7.2)

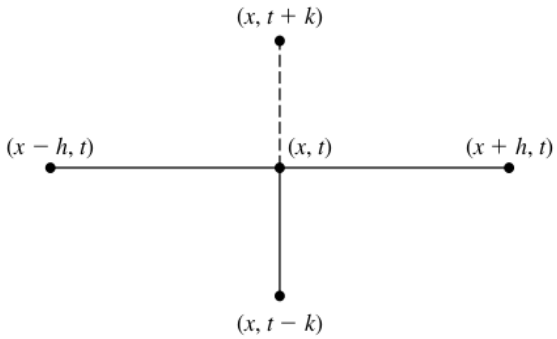
$$\frac{1}{h^2}\left[u\left(x+h,t\right)-2u\left(x,t\right)+u\left(x-h,t\right)\right]=\frac{1}{k^2}\left[u\left(x,t+k\right)-2u\left(x,t\right)+u\left(x,t-k\right)\right],$$

que puede también ser reordenada como

$$u\left(x,t+k\right)=\rho u\left(x+h,t\right)+2\left(1-\rho\right)u\left(x,t\right)+\rho u\left(x-h,t\right)-u\left(x,t-k\right) \tag{7.5}$$

donde $\rho=\frac{k^2}{h^2}$.

La figura siguiente muestra el punto $(x,t+k)$ y los puntos cercanos que entran en la ecuación (7.5).



Las condiciones de frontera en el problema (7.3) pueden ser escritas como

$$\begin{cases} u(x,0) & = & f(x), \\ \frac{1}{k}\left[u(x,k) - u(x,0)\right] & = & 0, \\ u(0,t) = u(1,t) & = & 0. \end{cases} \tag{7.6}$$

El problema definido por las ecuaciones (7.5) y (7.6) puede ser resuelto comenzando en la línea de $t = 0$, donde u es conocida, y luego avanzar una línea a la vez con $t = k$, $t = 2k$, $t = 3k$, ... Note que de la ecuación (7.6), nuestra solución aproximada satisface

$$u(x,k) = u(x,0) = f(x). \tag{7.7}$$

El uso de la aproximación $O(k)$ para u_t conduce a una baja precisión en la solución computarizada al problema (7.3). Supongamos que hay una fila en la cuadrícula con puntos $(x, -k)$. Dejamos $t = 0$ en la ecuación (7.5), y tenemos

$$u(x,k) = \rho u(x+h,0) + 2(1-\rho)u(x,0) + \rho u(x-h,0) - u(x,-k).$$

Ahora la aproximación de la diferencia central

$$\frac{1}{2k}\left[u(x,k) - u(x,-k)\right] = 0,$$

para $u_t(x,0) = 0$ se puede utilizar para eliminar el punto de la cuadrícula ficticia $(x, -k)$. Así que en lugar de la ecuación (7.7), ponemos

$$u(x,k) = \frac{1}{2}\rho\left[f(x+h) + f(x-h)\right] + (1-\rho)f(x),$$

porque $u(x,0) = f(x)$.

Los valores de $u(x,nk)$, $n \geq 2$, pueden ahora ser calculados de la ecuación (7.5).

7.2.3. Pseudocódigo

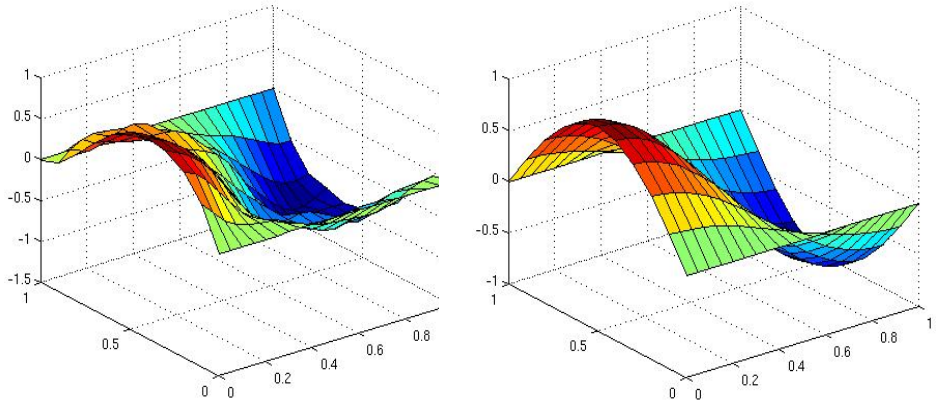
```
%solucion exacta
x=0:h:1; y=0:k:k*20;
[a b]=meshgrid(y,x);
u = sin(pi.*a).*cos(pi.*b)
surf(a, b, u)
```

.

```
n=10; m=20; h=0.1; k=0.05;
u=zeros(n+1,m+1); v=zeros(n+1,m+1);
ro=(k/h)^2;
for i=2:n
    x=i*h;
    w(i)=f(x);
    v(i)=(ro/2)*(f(x-h)+f(x+h))+(1-ro)*f(x);
end
for j=1:m+1
    for i=2:n
        u(i,j)=ro*(v(i+1)+v(i-1))+2*(1-ro)*v(i)-w(i);
    end
    for i=2:n
        w(i)=v(i);
        v(i)=u(i,j);
        t=j*k;
        x=i*h;
    end
end
disp(u)
x=0:h:1;
y=0:k:k*20;
[a b]=meshgrid(y,x);
surf(a, b, u)
```

Este pseudocódigo requiere el acompañamiento de funciones para calcular los valores de $f(x)$ y de la solución exacta. Nosotros elegimos $f(x) = \sin(\pi x)$ en el ejemplo. Asumimos que x esta en el intervalo $[0, 1]$, pero cuando h o n cambien, el intervalo puede ser $[0, b]$, esto es, $nh = b$. La solución numérica se calcula en las t lineas que corresponden a $1k$, $2k, \cdots$, mk.

A continuación se presentan las gráficas de la solución aproximada y la solución exacta



Los tratamientos más avanzados muestran que las proporciones $\rho = \frac{k^2}{h^2}$ no debe exceder a 1 si la solución de las ecuaciones en diferencias finitas converge a una solución del problema diferencial cuando $k \to 0$ y $h \to 0$. Por otra parte, si $\rho > 1$, los errores de redondeo que se producen en una etapa del computo probablemente se magnifiquen en etapas posteriores y por lo tanto arruinarían la solución numérica.

7.3. Ecuación de advección

Nos enfocamos ahora en la ecuación de advección (advección es la variación de un escalar en un punto dado por efecto de un campo vectorial)

$$\frac{\partial u}{\partial t} = -c\frac{\partial u}{\partial x},$$

aquí, $u = u(x,t)$ y $c = c(x,t)$ en el que se puede considerar a x como el espacio y t el tiempo. La ecuación de advección es una ecuación diferencial parcial hiperbólica que gobierna el movimiento de un escalar conservado, que es advectado por un campo de velocidad conocida. Por ejemplo, la ecuación de advección se aplica al transporte de sal disuelta en el agua. Incluso en una dimensión espacial y la velocidad constante, el sistema sigue siendo difícil de resolver.

Dado que la ecuación de advección es difícil de resolver numéricamente, el interés general se centra en soluciones de choque discontinuas, que son notoriamente difíciles para manejar esquemas numéricos.

Usando la aproximación por diferencias hacia adelante en el tiempo y las diferencias centrales para las aproximaciones en el espacio, tenemos

$$\frac{1}{k}\left[u(x,t+k) - u(x,t)\right] = -c\frac{1}{2h}\left[u(x+h,t) - u(x-h,t)\right],$$

y esto da

$$u(x,t+k) = u(x,t) - \frac{1}{2}\sigma\left[u(x+h,t) - u(x-h,t)\right],$$

donde $\sigma = \frac{h}{k}c(x,t)$. Todas las soluciones numéricas crecen en magnitud por todo el tiempo con pasos de tamaño k.

7.3.1. Solución a la ecuación de advección

Aplicaremos el esquema de Lax-Friedrichs para resolver la ecuación de advección, este esquema ya aplicado a la ecuación es:

$$\frac{1}{k}\left[v_m^{n+1} - \frac{1}{2}\left(v_{m+1}^n + v_{m-1}^n\right)\right] + a\frac{1}{2h}\left[v_{m+1}^n - v_{m-1}^n\right] = 0.$$

Se obtiene al utilizar aproximaciones de segundo orden para la derivada en el tiempo y el espacio, así:

$$\begin{aligned}
\frac{\partial u}{\partial t}(nh,mk) &\approx \frac{1}{2k}\left[u((n+1)k,mh) - u((n-1)k,mh)\right], \\
\frac{\partial u}{\partial x}(nh,mk) &\approx \frac{1}{2h}\left[u(nk,(m+1)h) - u(nk,(m-1)h)\right].
\end{aligned}$$

Antes definimos una malla en el plano (t, x). Sean h y k dos números reales positivos, entonces la malla estará compuesta por los puntos $(t_n, x_m) = (nk, mh)$, donde n y m son enteros arbitrarios. El conjunto de puntos (t_n, x_m) para un valor fijo dado de n es llamado malla al nivel n.

La función $u(t_n, x_m)$ es el valor de la función $u(t, x)$ en el punto (t_n, x_m) de la malla. Mientras que denotamos por v_m^n el valor de la función v en el punto (t_n, x_m).

Note que hacemos esto para diferenciar el valor de la solución $u(t_n, x_m)$ de la ecuación diferencial de su correspondiente aproximación v_m^n.

El esquema puede expresarse en términos de v_m^{n+1} como una combinación lineal de valores de v en los niveles n y $n-1$, así

$$v_m^{n+1} = \alpha v_{m+1}^n + \beta v_{m-1}^n$$

donde $\alpha = \frac{1}{2}(1 - a\lambda)$, $\beta = \frac{1}{2}(1 + a\lambda)$, con $\lambda = k/h$.

Para completar el problema, supongamos que las condiciones iniciales de nuestro problema son

$$\begin{cases} u(x, 0) = (\cos(\pi x))^2 & \text{si } |x| \leq 0.5 \\ u(x, 0) = 0 & \text{si } |x| \leq 0.5 \end{cases}$$

con dominio en en espacio $[-1, 1]$ y con condición de borde $u(-1, t) = 0$.

Para resolver este problema, crearemos una función en MATLAB, en la que como parámetro de funciones anónimas le pasamos las que nos dan las condiciones iniciales y de frontera, las cuales son:

```
function [U] = cborde(t)
%Función para la condicón de borde U(-1,t)

n=length(t);
U=zeros(1,n);

end
```

```
function [U] = cinicial(x)
%Función para la condición inicial U(x,0)
n=length(x);
U=zeros(1,n);
for i=1:n
    if abs(x(i)) <= 0.5
        U(i)=(cos(pi*x(i)))^2;
    else
        U(i)=0;
    end
end
end
```

Los demás parámetros son: $lambda = k/h$

a, constante del medio.

M, Número particiones en el espacio

xl, es un vector cuyas componentes son el borde inicial y final respectivamente

tl, es un vector, cuyas componentes son el tiempo inicial y final respectivamente.

V, es el argumento de salida. Siendo V, una matriz, en donde $V(j, i)$ es la elongación para cada posición y tiempo respectivamente.

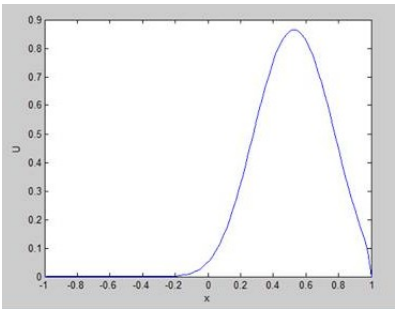
La función es la siguiente:

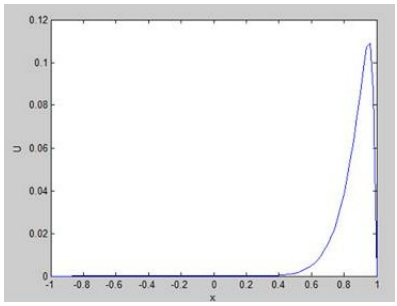
```
function [V] = laxfriedrichs(c_inicial,c_borde,lambda,a,M,xl,tl)
x=linspace(xl(1),xl(2),M);
h=(1/M);
k=lambda*h;

alfa=0.5*(1-(a*lambda));
beta=0.5*(1+(a*lambda));

t=tl(1):k:tl(2);
n=length(t);
V=zeros(M,n);


V(:,1)=c_inicial(x)';
V(1,:)=c_borde(t);

for ii=2:n
    for jj=2:M-1
        V(jj,ii)=alfa*V(jj+1,ii-1)+beta*V(jj-1,ii-1);
    end
end

%% Grafica de las soluciones
for ii=1:n
    plot(x',V(:,ii)), xlabel('x'), ylabel('U')
    pause(0.1)
end
end
```

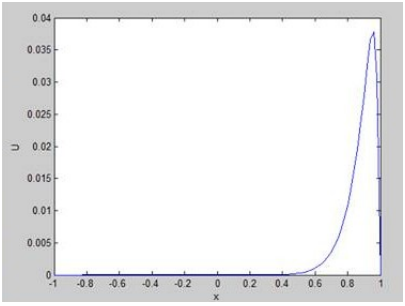
Ahora, para correr el código en la ventana de comandos de MATLAB escribimos lo siguiente:

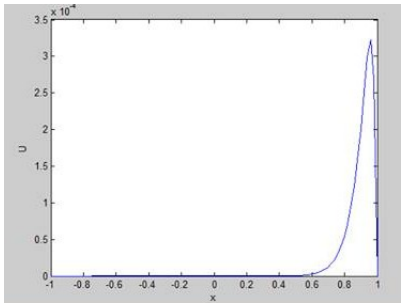
$$[V] = laxfriedrichs(@cinicial, @cborde, 0.5, 1, 100, [-1, 1], [0, 3])$$

La constante lambda, depende de cada esquema de diferencias le ponemos un valor de 0.5. En la parte final graficamos la aproximación, las gráficas que obtenemos son las elongaciones para cada instante de tiempo, es decir cada curva es la onda en un tiempo determinado. De ahí es que sale la animación. Si se grafican todas juntas, lo que se ve es la misma curva (mas o menos, dado que son aproximaciones) trasladada. Los tiempos los poder ver en el vector t. Algunas de las gráficas observadas son:









8. Capítulo 5. Problemas Elípticos

Una de las más importantes ecuaciones diferenciales parciales de la física matemática y la ingeniería es la ecuación de Laplace, que tiene la siguiente forma en dos variables:

$$\nabla^2 u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Estrechamente relacionado con esta, tenemos la ecuación de Poisson:

$$\nabla^2 u = g(x, y).$$

Estos son ejemplos de ecuaciones elípticas. Las condiciones de contorno asociados con ecuaciones elípticas generalmente difieren de aquellos para ecuaciones parabólicas e hiperbólicas. Consideramos aquí un problema modelo para ilustrar los procedimientos numéricos que se utilizan a menudo.

8.1. Problema Modelo de la ecuación de Helmholtz

Supongamos que una función $u = u(x, y)$ de dos variables es la solución a un cierto problema físico. Esta función es desconocida, pero tiene algunas propiedades que, en teoría, determinan su unicidad. Suponemos que en una región R dada en el plano xy,

$$\begin{cases} \nabla^2 u + fu = g \\ u\,(x, y) \ \text{ conocida en la frontera de } R. \end{cases} \tag{8.1}$$

Aquí, $f = f(x, y)$ y $g = g(x, y)$ se dan como funciones continuas definidas en R. La valores en la frontera podrían ser dados por una tercera función $u(x, y) = q(x, y)$ en el perímetro de R. Cuando f es una constante, esta ecuación diferencial parcial se llama la ecuación de Helmholtz. Se origina en la búsqueda de soluciones oscilatorias de las ecuaciones de onda.

8.2. Método de diferencias finitas

Como antes, nos encontramos con una solución aproximada de un problema tal por el método de diferencias finitas. El primer paso es seleccionar las fórmulas aproximadas para los derivadas de nuestro problema. En la situación actual, utilizamos la fórmula estándar

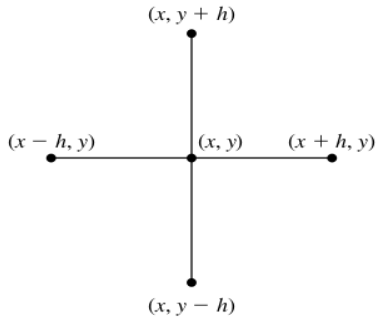
$$f''\,(x) \approx \frac{1}{h^2}\left[f\,(x + h) - 2f\,(x) + f\,(x - h)\right] \tag{8.2}$$

obtenida en una sección anterior. Si usamos una función de dos variables, obtenemos la fórmula

de 5 puntos para la aproximación de la ecuación de Laplace:

$$\nabla^2 u\,(x,y) \approx \frac{1}{h^2}\left[u\,(x+h,y)+u\,(x-h,y)+u\,(x,y+h)+u\,(x,y-h)-4u\,(x,y)\right] \tag{8.3}$$

Esta fórmula involucra los 5 puntos mostrados en la figura



El error local inherente en la fórmula de 5 puntos es

$$-\frac{h^2}{12}\left[\frac{\partial^4 u}{\partial x^4}\,(\xi,y)+\frac{\partial^4 u}{\partial x^4}\,(x,\eta)\right], \tag{8.4}$$

y por esta razón, la fórmula (8.3) se dice que proporciona una aproximación de orden $O(h^2)$. En otras palabras, si las mallas se utilizan con más y más pequeño espacios $h \to 0$, entonces el error que se comete en la sustitución $\nabla^2 u$ por su aproximación de diferencias finitas va a cero tan rápidamente como lo hace h^2. La ecuación (8.3) se llama la fórmula de cinco puntos porque se trata de valores de u en (x,y) y en los cuatro puntos mas cercanos de la malla.

Volviendo al problema modelo (8.1), cubrimos la región R por puntos de la malla

$$x_i = ih \quad y_j = jh \quad (i,j \geq 0). \tag{8.5}$$

Introducimos la notación abreviada:

$$u_{ij} = u(x_i,y_i), \;\; f_{ij} = f(x_i,y_i), \;\; g_{ij} = g(x_i,y_j). \tag{8.6}$$

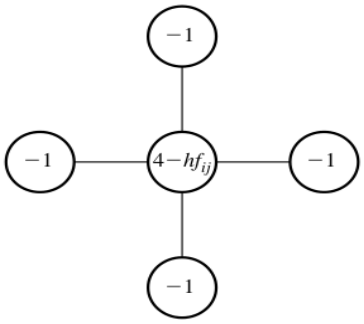
Con esto, la fórmula de cinco puntos adquiere una forma sencilla en el punto (x_i,y_j):

$$\left(\nabla^2 u\right)_{ij} \approx \frac{1}{h^2}\left[u_{i+1,j}+u_{i-1,j}+u_{i,j+1}+u_{i,j-1}+4u_{ij}\right]. \tag{8.7}$$

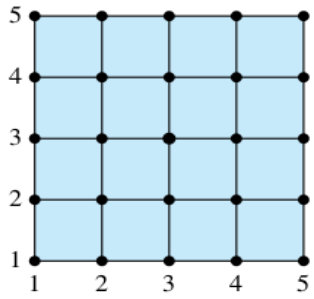
Si esta aproximación se hace en la ecuación diferencial (8.1), el resultado es

$$-u_{i+1,j}-u_{i-1,j}-u_{i,j+1}-u_{i,j-1}+\left(4-h^2 f_{ij}\right)u_{ij} = -h^2 g_{ij}. \tag{8.8}$$

Los coeficientes de esta ecuación se pueden ilustrar por una estrella de cinco puntos en la que cada punto se corresponde con el coeficiente de u en la malla



Para ser más específicos, se supone que la región R es el cuadrado unidad y que la malla tiene el espaciamiento $h = 1/4$



Obtenemos una ecuación lineal de la forma (8.8) para cada uno de los nueve puntos interiores de la malla. Estas nueve ecuaciones son las siguientes:

$$\begin{cases} -u_{21} - u_{01} - u_{12} - u_{10} + (4 - h^2 f_{11})\, u_{11} = -h^2 g_{11} \\ -u_{31} - u_{11} - u_{22} - u_{20} + (4 - h^2 f_{21})\, u_{21} = -h^2 g_{21} \\ -u_{41} - u_{21} - u_{32} - u_{30} + (4 - h^2 f_{31})\, u_{31} = -h^2 g_{31} \\ -u_{22} - u_{02} - u_{13} - u_{11} + (4 - h^2 f_{12})\, u_{12} = -h^2 g_{12} \\ -u_{32} - u_{12} - u_{23} - u_{21} + (4 - h^2 f_{22})\, u_{22} = -h^2 g_{22} \\ -u_{42} - u_{22} - u_{33} - u_{31} + (4 - h^2 f_{32})\, u_{32} = -h^2 g_{32} \\ -u_{23} - u_{03} - u_{14} - u_{12} + (4 - h^2 f_{13})\, u_{13} = -h^2 g_{13} \\ -u_{33} - u_{13} - u_{24} - u_{22} + (4 - h^2 f_{23})\, u_{23} = -h^2 g_{23} \\ -u_{43} - u_{23} - u_{34} - u_{32} + (4 - h^2 f_{33})\, u_{33} = -h^2 g_{33} \end{cases}$$

Este sistema de ecuaciones puede ser resuelto a través de la eliminación de Gauss, pero examinando más de cerca. Hay 45 coeficientes. Como u es conocido en los puntos de frontera, nos movemos estos 12 términos a la derecha, dejando sólo 33 elementos distintos de cero de los 81 en nuestro sistema de 9×9. La eliminación de Gauss estándar hace que una gran cantidad de

relleno, en la eliminación hacia adelante, es decir, las entradas que son cero son sustituidos por valores distintos de cero. Por lo tanto, buscamos un método que conserve la estructura dispersa de este sistema. Para ilustrar este sistema de ecuaciones es, lo escribimos en notación matricial:

$$Au = b. \tag{8.9}$$

Supongamos que ordenamos las incógnitas de izquierda a derecha y de abajo a arriba:

$$u = [u_{11}, u_{21}, u_{31}, u_{12}, u_{22}, u_{32}, u_{13}, u_{23}, u_{33}]^T. \tag{8.10}$$

Esto se denomina el orden natural. Ahora, la matriz de coeficientes es:

$$\begin{bmatrix} 4-h^2f_{11} & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4-h^2f_{21} & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4-h^2f_{31} & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4-h^2f_{12} & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4-h^2f_{22} & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4-h^2f_{32} & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4-h^2f_{13} & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4-h^2f_{23} & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4-h^2f_{33} \end{bmatrix}$$

y el lado derecho es

$$b = \begin{bmatrix} -h^2g_{11} + u_{10} + u_{01} \\ -h^2g_{21} + u_{20} \\ -h^2h_{31} + u_{30} + u_{41} \\ -h^2g_{12} + u_{02} \\ -h^2g_{22} \\ -h^2g_{32} + u_{42} \\ -h^2g_{13} + u_{14} + u_{03} \\ -h^2g_{23} + u_{24} \\ -h^2g_{33} + u_{34} + u_{43} \end{bmatrix}$$

Note que si $f(x, y) < 0$, entonces A es una matriz diagonal dominante.

8.3. Método Iterativo de Gauss-Seidel

Como las ecuaciones son similares en forma, se usan métodos iterativos para resolver estos sistemas, donde la matriz de coeficientes es dispersa. Despejando la incógnita diagonal, tenemos la ecuación (8.8) el método de Gauss-Seidel o iteración dada por

$$u_{ij}^{(k+1)} = \frac{1}{4-h^2f_{ij}} \left(u_{i+1,j}^{(k)} + u_{i-1,j}^{(k+1)} + u_{i,j+1}^{(k)} + u_{i,j-1}^{(k+1)} + h^2g_{ij} \right).$$

Si tenemos valores aproximados de las incógnitas en cada punto de la cuadrícula, esta ecuación puede ser utilizada para generar nuevos valores. Llamamos $u^{(k)}$ los valores actuales de las incógnitas en la iteración k y $u^{(k+1)}$ el valor en la siguiente iteración. Por otra parte, los nuevos valores a utilizar están disponibles.

El pseudocódigo del método de Gauss-Seidel en un rectángulo, es como sigue:

```
%gauss seidel en el cuadrado
function m= seidel(ax,ay,nx,ny,h,itmax,u)

for k=1:itmax
    for j=2:ny
        y=ay+j*h;
        for i=2:nx
            x=ax+i*h;
            v= u(i+1,j)+u(i-1,j)+u(i,j+1)+u(i,j-1);
            u(i,j)=(v-(h^2)*g(x,y))/(4-(h^2)*f(x,y));
        end
    end
end
m=u;
end
```

En esta función, escrita en Octave/MATLAB, se debe decidir el número de iteradas a calcular, $itmax$. Las coordenadas de la esquina inferior izquierda del rectángulo (a_x, a_y) y el tamaño de paso h específico. El número de puntos en x en la malla n_x y el número de puntos en y es n_y. El método de Gauss Seidel resuelve un sistema de ecuaciones de manera iterada.

8.4. Ejemplo Numérico y Pseudocódigo

Ilustremos este procedimiento en el problema de valores en la frontera

$$\begin{cases} \nabla^2 u - \frac{1}{25}u = 0 & \text{dentro de } R \text{ el cuadrado unidad} \\ u = q & \text{en la frontera de } R \end{cases} \tag{8.11}$$

donde $q = \cosh\left(\frac{1}{5}x\right) + \cosh\left(\frac{1}{5}y\right)$. Este problema tiene una solución conocida $u = q$. A continuación presentamos un código para el procedimiento de Gauss-Seidel, comenzando con $u = 1$ y teniendo 20 iteraciones. Notemos que sólo se necesitan 81 espacios de almacenamiento para la matriz en la solución del sistema lineal 49×49 de manera iterativa. Aquí, $h = 1/8$.

```
%solucion exacta
a=0:1/8:1; b=0:1/8:1; [x,y]=meshgrid(a,b);
z=cosh(0.2.*x)+cosh(0.2.*y)
surf(x,y,z)
```

```
%eliptico
nx=8; ny=8; itmax=50;
ax=0; bx=1; ay=0; by=1;
h=(bx-ax)/nx;

u=zeros(nx+1,ny+1);

for j=0:ny
    y=ax+j*h;
    u(1,j+1)= bandy(ax,y);
    u(nx+1,j+1)=bandy(bx,y);
end
for j=0:nx
    x=ay+j*h;
    u(j+1,1)= bandy(x,ay);
    u(j+1,ny+1)=bandy(x,by);
end
for j=2:ny
    y=ay+j*h;
    for i=2:nx
        x=ax+i*h;
        u(i,j)=ustar(x,y);
    end
end

m=seidel(ax,ay,nx,ny,h,itmax,u);
disp(m)

%graficamos la solucion aproximada
a=0:1/8:1; b=0:1/8:1;
[x,y]=meshgrid(a,b);
surf(x,y,m)
```

Para este problema modelo, se acompaña de las funciones

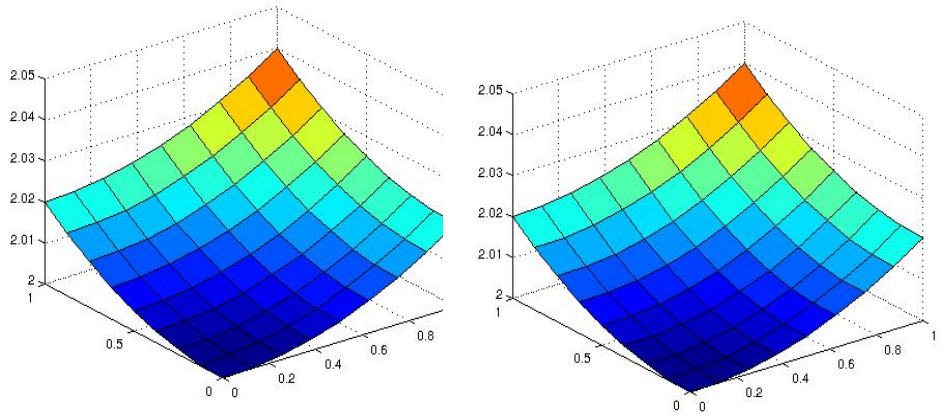
```
%funcion real
function v=bandy(x,y)
v=cosh(0.2*x)+cosh(0.2*y);
end
```

```
%funcion real
function a=f(x,y)
a=-0.04;
end
```

```
function m=g(x,y)
m=0;
end
```

```
%funcion real
function s=ustar(x,y)
s=1;
end
```

Graficando las soluciones exacta y la solución aproximada



Este ejemplo es una buena ilustración del hecho de que el problema numérico está resolviendo es el sistema de ecuaciones lineales (8.9), que es una aproximación discreta al problema con valor de frontera continuo (8.11). Al comparar la verdadera solución de (8.11) con la solución calculada del sistema, recordar que la discretización involucrado un error en la toma de la aproximación. Este error es $O(h^2)$. Con h tan grande como $h = 1/8$, la mayoría de los errores en la solución computarizada son debido al error de discretización.

Para obtener un mejor acuerdo entre los problemas discretos y continuos, se selecciona un tamaño para la malla mucho más pequeño. Por supuesto, el sistema lineal resultante tendrá una matriz de coeficientes que es extremadamente grande y bastante dispersa. Los métodos iterativos son ideales para la resolución de este tipo de sistemas que surgen de las ecuaciones en derivadas parciales.

9. Capítulo 6. Método de elementos finitos en una dimensión.

9.1. Problema modelo

El método de elementos finitos es una técnica general para construir soluciones aproximadas a problemas de valor en la frontera. El método implica dividir el dominio de la solución en un número finito de subdominios simples, los elementos finitos, y usando conceptos variacionales para construir una aproximación de la solución sobre la colección de elementos finitos. Debido a la generalidad y riqueza de las ideas que subyacen en el método se ha utilizado con notable éxito en la resolución de un amplio rango de problemas en prácticamente todas las áreas de ingeniería y matemática física.

Nuestro objetivo en este capítulo es dar una breve introducción a muchas ideas fundamentales las cuales forman la base del método. Para este propósito, confinamos nuestra atención a un ejemplo simple: En una dimensión, un problema de valor de frontera caracterizado por una ecuación diferencial lineal ordinaria de segundo orden, junto con un par de condiciones de frontera. Nos referiremos a este ejemplo como nuestro "problema modelo". Aunque el problema modelo no es ni difícil ni de muchas prácticas interesantes, tanto su estructura matemática y nuestro enfoque en la formulación de aproximación por elementos finitos son esencialmente el mismo en problemas más complejos de mayor significancia.

9.2. La formulación del problema modelo

Comenzamos por considerar, el encontrar una función $u = u(x), 0 \leq x \leq 1$, la cual satisface la siguiente ecuación diferencial y condiciones de frontera

$$\begin{cases} -u'' + u = x, & 0 < x < 1, \\ u(0) = 0, & u(1) = 0. \end{cases} \tag{9.1}$$

Un problema como este puede surgir en el estudio de la deflección de una cadena en una base elástica o en la distribución de la temperatura en una barra.

Los datos del problema consisten de la información dada:

El dominio de la solución (en este caso el dominio es el intervalo unitario $0 \leq x \leq 1$), la parte "no homogénea" de la ecuación diferencial (representada por la función $f(x) = x$ del lado derecho), los coeficientes de las derivadas de u (en este caso estas constantes son -1 y 1) y los valores de frontera que la solución satisface (en este caso, cero en $x = 0$ y $x = 1$).

Los datos de nuestro problema modelos son "suaves"; por ejemplo, el lado derecho $f(x) = x$ y los coeficientes son infinitamente diferenciables. Como una consecuencia de la suavidad,

existe una única función u la cual satisface la ecuación diferencial en cada punto del dominio así como las condiciones de frontera. En este ejemplo particular, la solución de (9.1) es $u(x) = x - (\sinh(x) / \sinh(1))$. Sin embargo, en aplicaciones más técnicas, una o ambas de estas características del problema se pierden o bien no hay solución para la formulación clásica del problema porque alguno de los datos no son suaves, o si la solución existe, no puede ser encontrada por la complejidad del dominio, los coeficientes o las condiciones de frontera.

Como un ejemplo de la primera clase de dificultad, supongamos que en lugar de $f(x) = x$, se da como parte de los datos del lado derecho de (9.1), para tener el problema

$$-u'' + u = \delta\left(x - \frac{1}{2}\right), 0 < x < 1;\ u(0) = 0 = u(1), \tag{9.2}$$

donde $\delta\left(x - \frac{1}{2}\right)$ es la delta de Dirac: el impulso unitario o "punto origen" concentrado en $x = \frac{1}{2}$. El hecho es que $\delta\left(x - \frac{1}{2}\right)$ no es siquiera una función pero es más bien una manera simbólica de describir operaciones en funciones suaves definidas por:[1]

$$\delta\left(x - \frac{1}{2}\right)\phi(x) = \phi\left(\frac{1}{2}\right)$$

para cualquier función suave ϕ que satisfaga las condiciones de frontera. Cualquier función u que satisface (9.2), debe tener una discontinuidad en su primera derivada u' en $x = \frac{1}{2}$; su segunda derivada u'' no existe (en el sentido tradicional) en $x = \frac{1}{2}$.

¡Algo parece estar mal!. ¿Cómo puede una función u satisfacer (9.2) en todas partes en el intervalo $0 < x < 1$ cuando su segunda derivada no puede existir en $x = \frac{1}{2}$ porque los datos dados de problema son irregulares?.

La dificultad es que nuestro requerimiento que una solución u para (9.2) satisfaga la ecuación diferencial en todo punto x, $0 < x < 1$, es también fuerte. Para superar esta dificultad, debemos reformular el problema de valor de frontera en una manera que admita condiciones mas débiles en la solución y sus derivadas.

Tales formulaciones son llamada formulaciones variacionales o débiles del problema y son diseñados para acomodar datos irregulares y soluciones irregulares, tal como estas en el problema (9.2), asi como soluciones muy suaves, tal como nuestro problema modelo (9.1).

Siempre que exista una solución suave para el problema clásico, también existe solución para el problema débil. Así, no perdemos nada por reformular un problema en manera débil, y ganamos significantes ventajas pudiendo considerar problemas con soluciones bastante irregulares. Más importante, la formulación débil del problema de valor de frontera es precisamente la formulación que usamos para construir la aproximación por elementos finitos de las soluciones.

[1] La operación $\delta\left(x - \frac{1}{2}\right)\phi$ es escrita a veces como $\int_0^1 \delta\left(x - \frac{1}{2}\right)\phi(x)\,dx = \phi\left(\frac{1}{2}\right)$ para todas las funciones infinitamente diferenciables que satisfacen las condiciones de frontera $\phi(0) = 0 = \phi(1)$. Pero incluso esto es incorrecto o, a lo mejor, simbólicamente, porque no existe función integrable que pueda producir esta acción dada una función suave ϕ.

9.3. Formulación variacional del problema

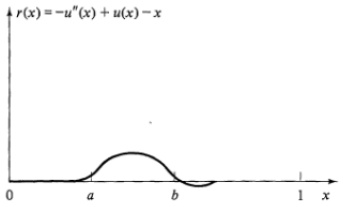
Una formulación débil del problema modelo (9.1) se da como sigue: Encontrar la función u tal que la ecuación diferencial, junto con las condiciones de frontera, sean satisfechas en el sentido de promedios ponderados. Por satisfacer en promedios ponderados la ecuación diferencial, significa que se requiere que:

$$\int_0^1 \left(-u'' + u\right) v dx = \int_0^1 x v dx, \tag{9.3}$$

para todos los miembros v de una clase adecuada de funciones. En (9.3) la función peso, o la función de prueba v, es cualquier función de x que es suave, para que la integral tenga sentido.[2] Con el fin de describir esta formulación débil del problema de manera mas concisa, introducimos la idea del conjunto de todas las funciones que son suaves para ser consideradas como funciones de prueba. Denotamos el conjunto de tales funciones, las cuales tienen valor cero en $x = 0$ y $x = 1$, por el símbolo H. La formulación variacional (9.3) de nuestro problema asume ahora la forma más compacta: Encontrar u tal que:

$$\begin{cases} \int_0^1 \left(-u'' + u - x\right) v dx = 0 \quad \text{para toda } v \in H \\ u\left(0\right) = 0 = u\left(1\right). \end{cases} \tag{9.4}$$

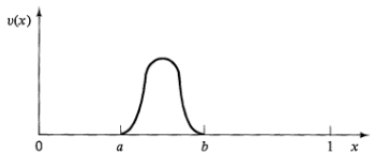
Después de reflexionar, si (9.4) es verdadero, no puede haber un subintervalo de longitud finita, por pequeño que sea, del intervalo $0 < x < 1$ dentro del cual la ecuación diferencial (9.1) falle de ser satisfecha en el sentido de promedio ponderado. Para ver esto, necesitamos solo hipotetizar la existencia de tal región y mostrar que como una consecuencia, (9.4) no sería satisfecha. Consideremos el residuo, o error, en la ecuación diferencial, definida por la función $r\left(x\right) = -u'' + u - x$. Supongamos que $r\left(x\right)$ es diferente de cero en alguna región pequeña, tal como en la figura



Correspondiente para esta particular $r\left(x\right)$ podemos elegir $v\left(x\right)$ como en la figura [3]

[2] Es fácil encontrar funciones que no son suaves para servir como funciones de prueba. Por ejemplo, si $u\left(x\right) = x - \sinh\left(x\right)$ y $v\left(x\right) = x^{-3}$, entonces ni $\int_0^1 \left(-u'' + u\right) v dx$ ni $\int_0^1 x v dx$ tienen valores finitos y (9.3) no tiene sentido. Hay, sin embargo, una multitud de funciones las cuales son perfectamente aceptables como funciones de prueba.

[3] Aunque no nos dieron la ecuación de $v\left(x\right)$, es claro que v es suave y suficiente para servir como una función de prueba.



Notemos que el integrando de (9.4) es positivo en el intervalo $a < x < b$ y cero en otra parte, vemos que la integral en (9.4) no puede eliminarse (i.e. (9.4) no es satisfecha). Así que u no puede ser una solución del problema (9.1). Mediante varias elecciones de v podemos "probar" la ecuación diferencial en todas las regiones de interés, así (9.4) de hecho requiere que (9.1) sea verdadero, en el sentido promedio ponderado, sobre cada subregión.

Esta formulación débil de nuestro problema, aunque aparentemente menos directa que la formulación clásica (9.1), tiene un cierto atractivo para los motivados por el argumento físico. En modelación de fenómenos físicos, a menudo es deseable la medida (o al menos para considerar la medición) de los datos y/o la solución del problema de valor de frontera. Ya que cualquier dispositivo de medición real (calibrador de tensión, par termoeléctrico, etc), tiene tamaño finito, estas cantidades pueden, a lo mejor, ser determinadas en el sentido promedio sobre pequeñas regiones y no sobre cualquier punto. La formulación débil del problema puede ser interpretada como asegurándonos que la solución puede parecer ser la correcta cuando probamos en cualquier lugar de la región.

9.3.1. Formulación variacional simétrica

La formulación débil (9.4) es tan válida y significativa como la formulación original (9.1), en efecto, la solución de (9.1) también satisface (9.4), de hecho es la (única) solución de (9.4).

Obtenemos una formulación débil alternativa, simétrica de (9.1), observando que, si u y v son suaves, entonces por la fórmula estándar de integración por partes tenemos:

$$\int_0^1 -u''v\,dx = \int_0^1 u'v'dx - u'v\,|_0^1\,.$$

Continuamos exigiendo que las funciones de prueba se hagan cero en los extremos, entonces $\int_0^1 -u''v\,dx = \int_0^1 u'v'dx$ para todas las funciones de prueba v, y por lo tanto (9.4) puede ser reemplazado por el siguiente problema variacional alternativo:

Encontrar $u \in H_0^1$ tal que

$$\int_0^1 (u'v' + uv - xv)\,dx = 0, \text{ para todo } v \in H_0^1. \tag{9.5}$$

Donde H_0^1 es una nueva clase de funciones. [4]

[4]Usaremos la notación $H_0^1(0,1)$ para describir esta clase de funciones, el superíndice "1" significa que los

Por otra parte, una vez más, cualquier solución de nuestro problema modelo (9.1) satisface (9.5), por lo que todavía no hemos perdido nada en esta reformulación. Sin embargo, ya que (9.4) contiene la segunda derivada de la solución u mientras que (9.5) tiene solo la primera derivada, vemos que pasando de (9.1) a (9.4) y de (9.4) a (9.5) hemos debilitado progresivamente la suavidad que se requiere para nuestra solución y así, progresivamente ampliamos la clase de datos para los cuales esta formulación del problema tiene sentido. Nos referiremos a la particular formulación débil definida en (9.5) como un problema variacional de valor de frontera.

Definiremos el conjunto de funciones admisibles H_0^1 como el conjunto de todas las funciones que se hacen cero en los puntos finales y cuya primera derivada es cuadrado integrable. Así, una función w es miembro de H_0^1 si

$$\int_0^1 (w')^2 \, dx < \infty, \text{ y } w(0) = 0 = w(1). \tag{9.6}$$

A pesar que hemos derivado la formulación variacional (9.5) de (9.1), es importante considerar que lo contrario es cierto: Consideraremos que (9.5) es el problema modelo de valor de frontera dado que deseamos resolver en lugar de (9.1). Habiendo resuelto (9.5), podemos preguntarnos si la solución es suave también es una solución "clásica", que es, una función que satisface (9.1), en cada x en $0 < x < 1$. Claramente, este punto de vista hará que sea posible que consideremos una amplia clase de problemas variacionales de valor de frontera los cuales no tienen solución clásica como formulación variacional equivalente de buenos problemas con solución clásica.

9.4. Aproximaciones de Galerkin

Consideramos el problema modelo con la siguiente forma variacional: Encontrar $u \in H_0^1$ tal que

$$\int_0^1 (u'v' + uv) \, dx = \int_0^1 xv \, dx, \text{ para todo } v \in H_0^1. \tag{9.7}$$

Ahora comenzamos con la pregunta de determinar la solución aproximada para (9.7), (y por lo tanto de (9.1)), y centramos nuestro enfoque en las propiedades de la clase H_0^1 de las funciones admisibles definidas en (9.6).

Hay dos propiedades fundamentales de H_0^1 además de las enumeradas en (9.6), que juegan un papel crucial en el tipo de aproximación que queremos hacer. Primero, H_0^1 es un espacio lineal de funciones, y segundo, es infinito dimensional.

Un "espacio lineal", significa que las combinaciones lineales de funciones en H_0^1 también son miembros de H_0^1. En otras palabras, si v_1 y v_2 son funciones de prueba arbitrarias y α y β son constantes arbitrarias, entonces $\alpha v_1 + \beta v_2$ es también una función de prueba.

Por "infinito dimensional" significa que es necesario especificar una infinidad de parámetros para definir de forma única una función de prueba v en el espacio. En efecto, si introducimos el

miembros v de esta clase de funciones tienen derivadas de orden 1, las cuales son cuadrado integrable en el intervalo $0 < x < 1$ y el subíndice "0" indica que $v = 0$ en $x = 0$ y $x = 1$.

conjunto de funciones

$$\psi_n(x) = \sqrt{2}\sin n\pi x, \quad n = 1, 2, 3, ... \tag{9.8}$$

y v es una función de prueba suave en H_0^1, entonces v se puede representar en la forma

$$v(x) = \sum_{n=1}^{\infty} a_n \psi_n(x), \tag{9.9}$$

donde los coeficientes escalares a_n están dados por

$$a_n = \int_0^1 v(x)\,\psi_n(x)\,dx. \tag{9.10}$$

Así, en vista de (9.9), una infinidad de coeficientes a_n pueden ser especificados con el fin de definir cualquier función $v \in H_0^1$; el espacio H_0^1 de funciones admisibles es, por lo tanto, infinito dimensional.

Supongamos que nos dan un conjunto infinito de funciones $\{\phi_1(x), \phi_2(x), \phi_3(x), ...\}$ en H_0^1 las cuales tiene la propiedad que cada función de prueba v en H_0^1 puede ser representada como una combinación lineal de $\phi_i(x)$ como una serie del tipo (9.9). En el mejor de los casos, podríamos utilizar igualmente las funciones trigonométricas suaves ψ_n definidas en (9.8) para este propósito, pero queremos hacer énfasis que $\phi_i(x)$ no necesita ser trigonométrica pero puede ser una función menos suave. Nuestro requerimiento básico es que cada $v \in H_0^1$ sea representada como combinación lineal de tales funciones del tipo

$$v(x) = \sum_{i=1}^{\infty} \beta_i \phi_i(x), \tag{9.11}$$

donde las β_i son constantes y la serie converge en un sentido apropiado[5] para el espacio H_0^1. Un conjunto de funciones $\{\phi_i\}$ con estas propiedades se dice que proporciona una base para H_0^1 y las funciones ϕ_i son llamadas funciones base.

Está claro que si tomamos sólo un número finito N de términos en la serie (9.11), entonces obtendremos sólo una aproximación v_N de v:

$$v_N(x) = \sum_{i=1}^{N} \beta_i \phi_i(x). \tag{9.12}$$

Las N funciones base $\{\phi_1(x), \phi_2(x), ..., \phi_N(x)\}$ define un subespacio N-dimensional $H_0^{(N)}$ de H_0^1. El subespacio $H_0^{(N)}$ es de dimensión finita[6] N porque cada función v_N en $H_0^{(N)}$ está determinado por una combinación lineal de las N funciones $\phi_1, ..., \phi_N$ por (9.12). $H_0^{(N)}$ es un

[5] Si v_N está dada por (9.12), entonces v_N converge a la función v en sentido H_0^1 si

$$\lim_{N\to\infty} \int_0^1 \left[(v - v_N)^2 + (v' - v_N')^2\right] dx = 0$$

[6] Aquí asumimos que las N funciones $\phi_1, ..., \phi_N$ son linealmente independientes, es decir, es imposible encontrar N coeficientes $\beta_1, \beta_2, ..., \beta_N$ no todos iguales a cero, tal que $\sum_{n=1}^{N} \beta_i \phi_i(x) = 0$ para todo x.

subespacio de H_0^1 porque cada ϕ_i, $i = 1, 2, ..., N$, es, por definición, un miembro de H_0^1. Por ejemplo, $\{\phi_1, \phi_2, \phi_3\}$ es una base para un subespacio 3-dimensional $H_0^{(3)}$ de H_0^1; $\{\phi_1, \phi_2, \phi_3, \phi_4\}$ define un subespacio 4-dimensional de H_0^1; y etcétera.

Consideramos ahora el método de Galerkin para construir una solución aproximada para el problema variacional de valor de frontera (9.7). *El método de Galerkin consiste de buscar una solución aproximada a* (9.7) *en un subespacio finito-dimensional* $H_0^{(N)}$ *del espacio* H_0^1 *en lugar de en todo el espacio* H_0^1. Así, en lugar de abordar el problema infinito dimensional (9.7), buscamos una solución aproximada u_N en $H_0^{(N)}$ de la forma

$$u_N(x) = \sum_{i=1}^{N} \alpha_i \phi_i(x), \tag{9.13}$$

la cual satisface (9.7) con H_0^1 reemplazado por $H_0^{(N)}$. En otras palabras, la formulación variacional del problema aproximado es: Encontrar $u_N \in H_0^{(N)}$ tal que

$$\int_0^1 (u_N' v_N' + u_N v_N)\, dx = \int_0^1 x v_N\, dx, \text{ para todo } v_N \in H_0^{(N)}. \tag{9.14}$$

Como las ϕ_i son conocidas, u_N se determinará cuando los N coeficientes α_i en (9.13) sean determinados.

Ahora vamos a ver cómo podemos calcular los coeficientes α_i. Primero observamos que todas las funciones de prueba v_N son combinaciones lineales de las funciones base ϕ_i de la forma (9.12), los β_i son constantes arbitrarias. Notar de nuevo que v_N en (9.12) puede tomar los valores de cualquier función en $H_0^{(N)}$ a través de una elección adecuada de las constantes β_i.

Para determinar los valores específicos, α_i, de estos coeficientes que caracterizará la solución aproximada u_N, introducimos (9.12) y (9.13) en (9.14) para obtener la condición

$$\int_0^1 \left\{ \frac{d}{dx}\left[\sum_{i=1}^{N} \beta_i \phi_i(x)\right] \frac{d}{dx}\left[\sum_{j=1}^{N} \alpha_j \phi_j(x)\right] + \left[\sum_{i=1}^{N} \beta_i \phi_i(x)\right]\left[\sum_{j=1}^{N} \alpha_j \phi_j(x)\right] - x \sum_{i=1}^{N} \beta_i \phi_i(x) \right\} dx \tag{9.15}$$
$$= 0 \quad \text{para todo } \beta_i,\ i = 1, 2, ..., N$$

expandiendo (9.15) y factorizando los coeficientes β_i tenemos

$$\sum_{i=1}^{N} \beta_i \left(\sum_{j=1}^{N} \left\{ \int_0^1 \left[\phi_i'(x)\, \phi_j'(x) + \phi_i(x)\, \phi_j(x)\right] dx \right\} \alpha_j - \int_0^1 x \phi_i(x)\, dx \right) = 0 \tag{9.16}$$
$$\text{para todo } \beta_i,\ i = 1, 2, ..., N$$

donde $\phi_i'(x) = d\phi_i(x)/dx$.

La estructura de (9.16) se observa más fácilmente reescribiéndola en la forma más compacta

$$\sum_{i=1}^{N} \beta_i \left(\sum_{j=1}^{N} K_{ij} \alpha_j - F_i \right) = 0 \tag{9.17}$$

para todas las elecciones de β_i, donde

$$K_{ij} = \int_0^1 \left[\phi_i'(x)\,\phi_j'(x) + \phi_i(x)\,\phi_j(x)\right] dx \tag{9.18}$$

y

$$F_i = \int_0^1 x\phi_i(x)\,dx \tag{9.19}$$

y en el que $i, j = 1, 2, ..., N$.

La matriz $N \times N$ de números $\mathbf{K} = [K_{ij}]$ que usualmente se conoce como la matriz de rigidez; el vector columna $N \times 1$, $\mathbf{F} = \{F_i\}$ es conocido como el vector de peso. Como los ϕ_i son conocidos, los números K_{ij} y F_i pueden ser calculados directamente por las fórmulas (9.18) y (9.19).

Como los β_i son arbitrarios, (9.17) representa N ecuaciones que son satisfechas por los α_j, más bien que por la ecuación que puede parecer. Para ver esto, consideremos la siguiente elección natural para el conjunto de parámetros: $\beta_1 = 1, \beta_i = 0$ para $i \neq 1$. Entonces (9.17) produce

$$\sum_{j=1}^{N} K_{1j}\alpha_j = F_1$$

Siguiendo, el conjunto $\beta_2 = 1, \beta_i = 0$ para $i \neq 2$, de modo que

$$\sum_{j=1}^{N} K_{2j}\alpha_j = F_2$$

Continuando de esta manera, llegamos a un sistema de N ecuaciones lineales en los N coeficientes desconocidos α_j:

$$\sum_{j=1}^{N} K_{ij}\alpha_j = F_i, \;\; i = 1, 2, ..., N. \tag{9.20}$$

Como las funciones ϕ_i son independientes, las ecuaciones (9.20) serán independientes, y por lo tanto la matriz de riguidez $\mathbf{K}$ será invertible. Se sigue que los coeficientes α_j están únicamente determinados por (9.20) y son de la forma

$$\alpha_j = \sum_{i=1}^{N} \left(K^{-1}\right)_{ji} F_i, \tag{9.21}$$

donde $\left(K^{-1}\right)_{ji}$ son los elementos de la inversa de $\mathbf{K}$. La solución aproximada u_N es determinada ahora introduciendo (9.21) en (9.13).

Algunas de las razones por las que la formulación variacional simétrica (9.5) de nuestro problema modelo es preferible sobre la formulación débil (9.4) son:

Nuestra aproximación de la formulación simétrica ha dejado una matriz de rigidez simétrica en (9.18), mientras que una formulación asimétrica nunca lo hará. Esta simetría nos proporciona la oportunidad de reducir el esfuerzo computacional para obtener una solución aproximada.

Si se usa la formulación simétrica de nuestro problema modelo, el método de Galerkin proporciona la mejor aproximación posible de la solución u en $H_0^{(N)}$.

Para la formulación simétrica, el espacio de funciones de ensayo y de funciones de prueba coincide; por lo tanto, se necesita sólo un conjunto de funciones base ϕ_i para construir tales aproximaciones.

Es importante notar que la calidad de la aproximación esta completamente determinada por la elección de las funciones ϕ_i: una vez que estas han sido elegidas, la determinación de los coeficientes α_j se reduce a una cuestión de cálculo.

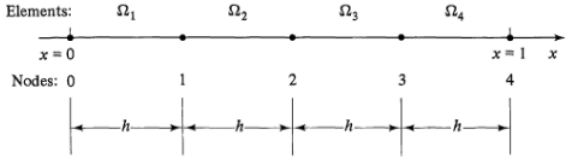
9.5. Funciones base

Mientras que el método de Galerkin proporciona una estrategia elegante para construir aproximaciones a la solución de un problema de valor de frontera, presenta un defecto: en el método que hemos descrito, no existe una manera sistemática de construir razonablemente las funciones base ϕ_i para la aproximación de las funciones de prueba v_N. Aparte de ser miembros independientes de H_0^1, son arbitrarias. El análisis se deja con un gran número de posibilidades a la disposición y con el conocimiento incómodo que la calidad de su solución aproximada podría depender muy fuertemente de las propiedades de las funciones base que se eligen. La situación es peor en problemas de valor de frontera en dos y tres dimensiones en el cual las funciones ϕ_i deben estar diseñados para ajustarse a las condiciones de frontera en dominios con geometría compleja.

Por otra parte, una mala elección de ϕ_i podría producir una matriz de rigidez mal condicionada así que el sistema lineal (9.20) podría ser difícil de resolver con un aceptable limite de exactitud. Por estas razones, el método clásico de Galerkin es de uso limitado. Estas dificultades pueden ser resueltas usando el método de elementos finitos.

El método de elementos finitos provee una técnica general y sistemática para construir funciones base para aproximaciones a la solución de problemas de valor de frontera. La idea principal es que las funciones base ϕ_i puedes ser definidas por piezas sobre subregiones del dominio llamadas elementos finitos y que sobre cualquier subdominio, las ϕ_i pueden ser elegidas para ser funciones muy simples tales como polinomios de grado pequeño.

Para construir tal conjunto de funciones base por trozos, primero particionamos el dominio (es decir, el intervalo $0 \leq x \leq 1$) de nuestro problema en un número finito de elementos.



La figura muestra, por ejemplo, el dominio de nuestro problema modelo particionado en cuatro elementos denotados Ω_i, $i = 1, 2, 3, 4$. Siguiendo una notación estándar, la longitud de cada elemento finito Ω_i será denotada h_i. Como los elementos en este ejemplo indicados en la figura

son de igual longitud, designaremos la longitud de los elementos en este caso por h.

Dentro de cada elemento, ciertos puntos se identifican, llamando nodos o puntos nodales, los cuales juegan un papel importante en la construcción de elementos finitos. En el ejemplo indicado en la figura anterior, los cinco nodos están tomados como los puntos en los extremos de cada elemento; estos están enumerados de 0 a 4 en la figura. A la colección de elementos y puntos nodales que componen el dominio del problema aproximado se le refiere en ocasiones como malla de elementos finitos.

Introducimos un leve cambio en la notación. En la sección precedente, denotamos las funciones de prueba y la solución por v_N y u_N, respectivamente, donde N era un parámetro que indicaba el número de funciones base usadas en la definición de $H_0^{(N)}$. Así, denotarmos, v_N, u_N y $H_0^{(N)}$ por v_h, u_h y H_0^h en las subsiguientes discusiones.

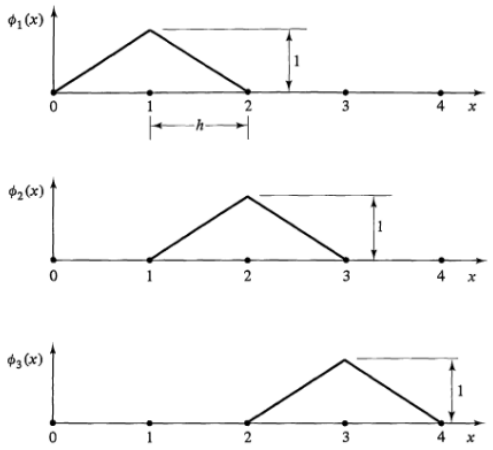
Teniendo construida una malla de elementos finitos para nuestro problema modelo (tal como en la figura), procedemos a construir el correspondiente conjunto de funciones base usando el siguiente criterio fundamental:

Las funciones base se generan por funciones simples definidas a trozos (elemento por elemento) sobre la malla de elementos finitos.

Las funciones base son suaves, siendo miembros de las clase H_0^1 de funciones de prueba.

Las funciones base se eligen de tal manera que los parámetros α_i que definen la solución aproximada $u_h (= u_N$, recordamos (9.13)) son precisamente los valores de $u_h(x)$ en los puntos nodales.

Un conjunto simple, pero perfectamente adecuado, de funciones base es el mostrado en la siguiente figura



Si las coordenadas de los nodos están denotadas por $x_i, (i = 0, 1, 2, 3, 4)$, entonces las funciones

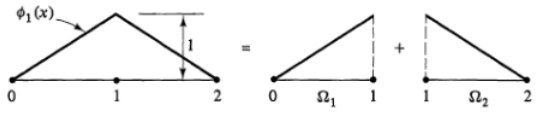
mostradas para $i = 1, 2, 3$ estan dadas por

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{h_i} & \text{para } x_{i-1} \leq x \leq x_i \\ \frac{x_{i+1} - x}{h_{i+1}} & \text{para } x_i \leq x \leq x_{i+1} \\ 0 & \text{para } x \leq x_{i-1} \text{ y } x \geq x_{i+1} \end{cases} \tag{9.22}$$

donde $h_i = x_i - x_{i-1}$ es la longitud del elemento Ω_i, su primera derivada es

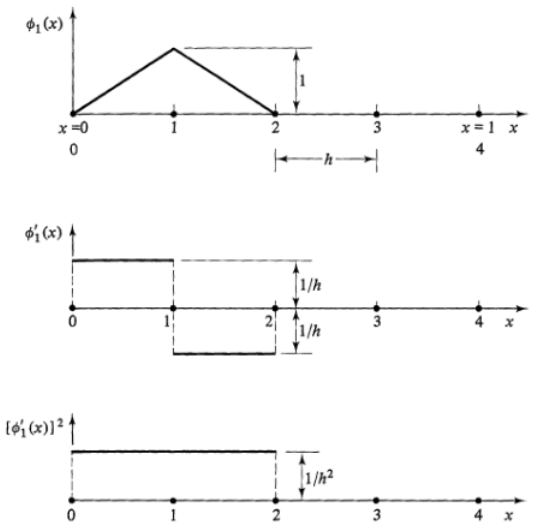
$$\phi_i'(x) = \begin{cases} \frac{1}{h_i} & \text{para } x_{i-1} < x < x_i \\ \frac{-1}{h_{i+1}} & \text{para } x_i < x < x_{i+1} \\ 0 & \text{para } x < x_{i-1} \text{ y } x > x_{i+1} \end{cases} \tag{9.23}$$

Para demostrar que estas funciones base satisfacen el criterio anterior, primero observamos que cada función ϕ_i, $i = 1, 2, 3$ es el resultado de poner juntas funciones lineales a trozos definidas sobre cada elemento finito. Por ejemplo, la función ϕ_1 asociada con el nodo 1 es producida combinando la función lineal definida en el elemento Ω_1 y la función lineal definida en el elemento Ω_2, como se muestra en la figura.



Esto ilustra lo que queremos decir cuando decimos que las funciones base son "generadas por funciones simples definidas a trozos (elemento por elemento) sobre la malla del elemento finito". La fuerza del método de elementos finitos descansa en la manera particular de construir las funciones base. Como un resultado de tal construcción, la aproximación (9.14) de nuestro problema puede ser formulada un elemento a la vez, la formulación final se obtiene mediante la suma de las contribuciones proporcionadas por cada elemento.

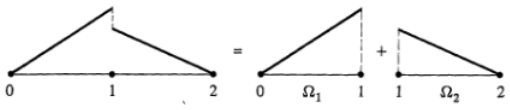
Volviendo ahora al criterio 2, recordemos que $\phi_i \in H_0^1$, $i = 1, 2, 3$ y cada una debe tener primera derivada cuadrado integrable y debe hacerse cero en $x = 0$ y $x = 1$. Las funciones mostradas en la figura claramente satisfacen la condición de frontera. ¿Son sus derivadas cuadrado integrable?, la respuesta es si, porque las derivadas de cada ϕ_i es una función escalonada, como las mostradas a continuación,



y además, $[\phi_i']^2$ es integrable. En efecto, la integral de $[\phi_1']^2$ es simplemente el área bajo la curva indicada en la figura anterior

$$\int_0^1 [\phi_1'(x)]^2 = \frac{1}{h^2} 2h = 2h^{-1} < \infty.$$

Lo crítico aquí es que al poner juntas las funciones lineales a trozos definidas razonablemente sobre los elementos para formar cada una de nuestras funciones base, la función adyacente coincida perfectamente en los nodos comunes. Entonces las $\phi_i(x)$ será continua a lo largo del domino del problema. Si este no es el caso, la función producida sufre un "salto" (una discontinuidad) en el nodo, tal como se indica en la figura



Tales funciones discontinuas no tendrán derivadas cuadrado integrable y por lo tanto, no pertenecerá a nuestra clase de funciones admisibles H_0^1.

Finalmente, llegamos al criterio 3: los parámetros α_i definiendo u_h deberían ser los valores de u_h en los puntos nodales. El criterio no es difícil de satisfacer si cada función base tiene la propiedad que su valor sea uno en un nodo y cero en los otros nodos. Específicamente, requerimos que si x_j es la coordenada en x del nodo j, entonces

$$\phi_i(x_j) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases} \tag{9.24}$$

En nuestro ejemplo, $i = 1, 2, 3$ y $j = 0, 1, 2, 3, 4$. Es claro que las funciones base lineales a trozos mostradas anteriormente satisfacen (9.24) (es decir, $\phi_1(x_1) = 1$, pero $\phi_j(x_j) = 0$ para

$j = 0, 2, 3, 4$), así que las funciones base definidas en (9.22) satisfacen el criterio 3. Note que $i = 0, 4$ no estan incluidos ya que las funciones base están obligadas a satisfacer las condiciones homogéneas en los extremos.

Sea v_h una función en H_0^h. De acuerdo con (9.13) y a nuestra nueva notación,

$$v_h(x) = \sum_{i=1}^{N} \beta_i \phi_i(x) \qquad (N = 3)$$

Sea v_j el valor de la función v_h en un punto nodal arbitrario j (es decir, $v_j = v_h(x_j)$) y manteniendo (9.24). Para nuestro ejemplo,

$$v_j = \sum_{i=1}^{3} \beta_i \phi_i(x_j) = \beta_j, \quad j = 1, 2, 3.$$

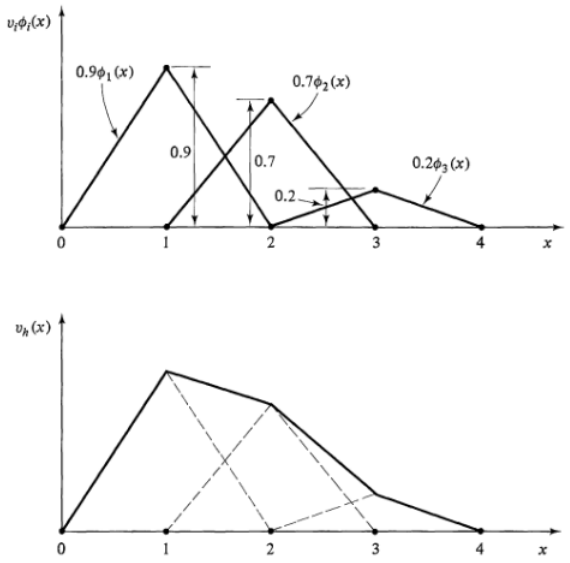
Se sigue que la representación en elementos finitos de v_h toma la forma

$$v_h(x) = \sum_{i=1}^{N} v_i \phi_i(x), \qquad v_i = v_h(x_i). \tag{9.25}$$

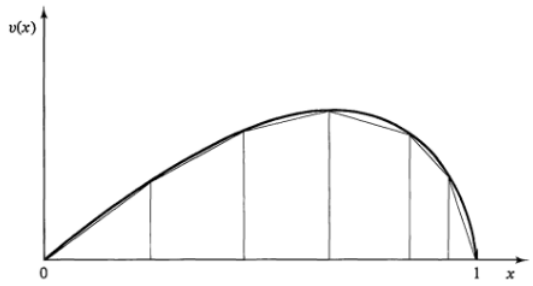
Es importante entender como los términos en (9.25) se suman para dar una representación continua de v_h. Por ejemplo, supongamos que los valores de v_h en los nodos $1, 2$ y 3 son $0.9, 0.7$ y 0.2 respectivamente. Entonces, sustituyendo estos valores en (9.25) da

$$v_h(x) = 0.9\phi_1(x) + 0.7\phi_2(x) + 0.2\phi_3(x),$$

de modo que estas tres componentes se combinan para dar una función continua a trozos, como la mostrada en la siguiente figura



Hay un punto final que se debe hacer aquí, que es sugerido por la forma de la función lineal a trozos v_h en la figura anterior. Supongamos que la solución actual de nuestro problema modelo es la función suave indicada en la figura.



Si consideramos la forma de esta función sobre un subintervalo suficientemente pequeño de su dominio, entonces es claro que es casi lineal sobre este intervalo, como el indicado en la figura. Si u es aproximada por funciones lineales a trozos con valores coincidiendo con los de u en los nodos, el resultado es una función poligonal muy parecida a u. Esto es una interpolación lineal a trozos de la solución exacta u. Cuando la malla se refina (es decir, cuando aumente en número de elementos y h disminuye), la interpolación de elementos finitos se convierte progresivamente cercana a u. Por otra parte, nuestra aproximación de elementos finitos u_h para la solución del problema de valor de frontera también será lineal a trozos pero sus valores nodales por lo general no están de acuerdo con los de la solución exacta. De la misma manera que la interpolación, la soluciones aproximadas u_h parecen tener la propiedad de producir progresivamente mejores aproximaciones a u cuando la malla es refinada. Estas ideas, por supuesto, está estrechamente relacionadas con los conceptos que subyacen del cálculo diferencial.

9.6. Cálculo de elementos finitos

Volviendo a la aproximación de Galerkin de el problema variacional de valor de frontera (9.7) usando la técnica de elementos finitos para construir las funciones base ϕ_i. El problema aproximado, entonces, consiste de encontrar $u_h \in H_0^h$, donde H_0^h es un subespacio de H_0^1 definido por las elecciones particulares de ϕ_i, tal que

$$\int_0^1 (u_h' v_h' + u_h v_h)\, dx = \int_0^1 x v_h\, dx, \text{ para todo } v_h \in H_0^h, \tag{9.26}$$

en el cual $u_h = \sum_{i=1}^N u_i \phi_i$ y u_i son los valores de u_h en el punto nodal de la malla de elementos finitos. En vista de (9.20), esto conduce al sistema lineal de ecuaciones

$$\sum_{j=1}^{N} K_{ij} u_j = F_i, \qquad i = 1, 2, ..., N \tag{9.27}$$

donde K_{ij} y F_i estan definidas por (9.18) y (9.19).

Examinamos a continuación algunas propiedades especiales e importantes de la matriz de rigidez $\mathbf{K}$ y del vector de carga $\mathbf{F}$.

1. Suma de rigideces: Esta es, quizás la propiedad más importante calculada usando elementos finitos. Supongamos que usamos, como un ejemplo, la malla de elementos finitos y las funciones base ϕ_1, ϕ_2, ϕ_3, indicadas antes, en vista de (9.18), cada entrada K_{ij}, se obtiene integrando $\left(\phi_i'\phi_j' + \phi_i\phi_j\right)$ sobre todo el dominio $0 \leq x \leq 1$. Pero la operación de integrar es aditiva (es decir, $\int_0^1 f\,dx = \int_0^{1/2} f\,dx + \int_{1/2}^1 f\,dx$), así que K_{ij} puede ser calculado como la suma

$$\begin{aligned} K_{ij} &= \int_0^1 \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx \\ &= \int_0^h \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx + \int_h^{2h} \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx \\ &\quad + \int_{2h}^{3h} \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx + \int_{3h}^1 \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx \\ &= \sum_{e=1}^4 \int_{\Omega_e} \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx \end{aligned} \tag{9.28}$$

donde $\int_{\Omega_e}$ denota la integración sobre el elemento Ω_e.

Sean los términos

$$K_{ij}^e = \int_{\Omega_e} \left(\phi_i'\phi_j' + \phi_i\phi_j\right)\,dx \tag{9.29}$$

que representan las componentes de la matriz de rigidez para el elemento finito Ω_e. Entonces,

$$K_{ij} = \sum_{e=1}^4 K_{ij}^e. \tag{9.30}$$

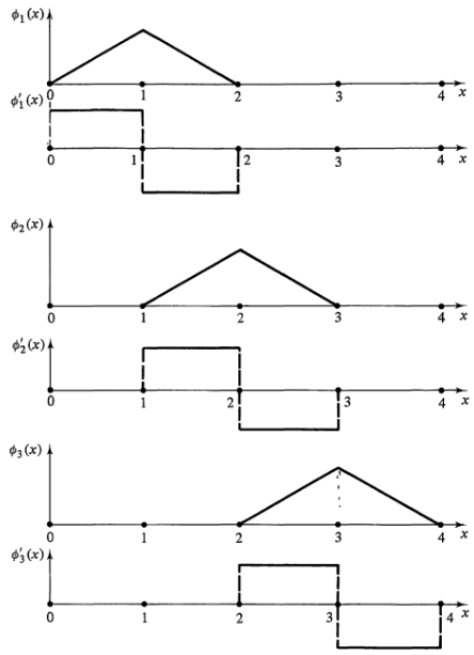
Similarmente,

$$F_i = \sum_{e=1}^4 F_i^e \quad \text{y} \quad F_i^e = \int_{\Omega_e} x\phi_i\,dx, \tag{9.31}$$

donde F_i^e son las componentes del vector de carga para el elemento finito Ω_e.

El hecho que K_{ij} y F_i puedan ser calculados como la suma de contribuciones de cada elemento es una característica clave del método de elementos finitos. Porque de esta propiedad elemental, es posible generar $\mathbf{K}$ y $\mathbf{F}$ calculando sólo los elementos de las matrices $\mathbf{K}^e$ y $\mathbf{F}^e$ para un elemento Ω_e y luego construir $\mathbf{K}$ y $\mathbf{F}$ como las sumas indicadas en (9.30) y (9.31).

2. Dispersidad de $\mathbf{K}$*:* Para nuestro problema modelo, con la malla indicada antes, debemos calcular nueve números, K_{ij}, $i, j = 1, 2, 3$, a fin de llegar a la matriz de rigidez de nuestra aproximación. Sin embargo, al examinar las gráficas:



revela que ϕ_1 y ϕ_1' son diferentes de cero sólo en los elementos Ω_1 y Ω_2 adyacentes al nodo 1. Similarmente, ϕ_2 y ϕ_2' no son cero sólo en los elementos Ω_2 y Ω_3 adyacentes al nodo 2, y ϕ_3 y ϕ_3' no son cero sólo en los elementos Ω_3 y Ω_4 adyacentes al nodo 3. Consecuentemente, los productos $\phi_i\phi_j$ y $\phi_i'\phi_j'$ no son cero sólo donde el soporte de las funciones base ϕ_i y ϕ_j se "superponen". Por ejemplo, los productos $\phi_1\phi_2$ y $\phi_1'\phi_2'$ no son cero sólo en el elemento 2, mientras que los productos $\phi_1\phi_3$ y $\phi_1'\phi_3'$ son cero en todas partes. Por lo tanto, las integrales K_{12}, K_{21} no son cero pero $K_{13} = K_{31} = 0$, automáticamente. Se sigue que si los nodos i y j no pertenecen al mismo elemento, entonces $K_{ij} = 0$. Esto implica que en una malla de muchos elementos, muchas de las entradas K_{ij} de la matriz serán cero. Las matrices con muchos ceros se llaman *dispersas* y en nuestra elección particular de funciones base de elemento finitos ϕ_i ha llevado a la dispersidad de $\mathbf{K}$.

La estructura final de la matriz de rigidez $\mathbf{K}$ es notable. Si numeramos los nodos secuencialmente, las entradas que no son cero aparecen agruparse cerca de la diagonal principal de la matriz. Fuera de esta "banda" de elementos, todas las entradas son cero.

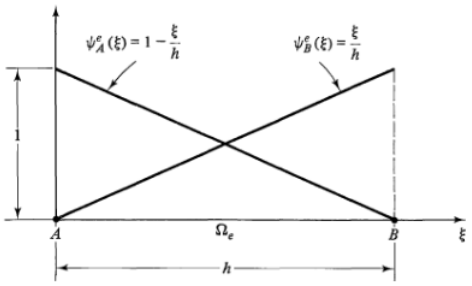
3. Simetría de $\mathbf{K}$*:* Intercambiando i y j en la expresión integral de K_{ij} no cambian el valor calculado, así que $K_{ij} = K_{ji}$ y la matriz de rigidez para el problema modelo es simétrica. No siempre tendremos simetría (considerar la contribución que surge si aparece la primera derivada u' en la ecuación diferencial original). Esta simetría de $\mathbf{K}$ no tiene nada que ver con la elección de funciones base y es totalmente dependiente de la forma del problema variacional a resolver.

Las propiedades de la matriz de rigidez descritas juegan un papel central en la estrategia de programar los cálculos de elementos finitos.

Volvamos ahora al problema de calcular realmente una solución aproximada de nuestro problema modelo en la malla dada. Usamos la suma de la propiedad 1 las contribuciones de la integral de

elemento individuales para la matriz de rigidez **K** y el vector de carga **F**. Debido a la simetría, las cálculos esenciales necesitan hacerse solo en un elemento finito Ω_e.

Empezamos el cálculo de los elementos de la matriz por considerar el elemento representativo Ω_e mostrado en la figura siguiente



e introducimos al nodo local los índices A y B. Sea ξ la coordenada local en este elemento representativo con su origen en el nodo izquierdo A de Ω_e. Entonces cuando x pasa de x_A a x_B, ξ va desde 0 hasta h. Tenemos, $\xi = x - x_A$.

Como las funciones base ϕ_i son construidas juntando polinomios definidos localmente sobre cada elemento. Nos referimos a estas partes componentes como funciones de forma del elemento. Por ejemplo, la función base ϕ_A en el nodo A en la malla se produce combinando funciones de forma del elemento definidas en los elementos conectados al nodo A.

Sean ψ_A^e y ψ_B^e las funciones de forma definidas para el elemento Ω_e, y mostradas en la figura anterior. Como estas son partes simples de ϕ_A y ϕ_B, estas funciones de forma estan dadas en términos de la coordenada local ξ por

$$\psi_A^e(\xi) = 1 - \frac{\xi}{h}, \qquad \psi_B^e(\xi) = \frac{\xi}{h}.$$

Claramente,

$$\psi_A^{e\prime}(\xi) = -\frac{1}{h}, \qquad \psi_B^{e\prime}(\xi) = \frac{1}{h}.$$

De acuerdo a (9.29), los elementos para la matriz de coeficientes de nuestro elemento Ω_e son

$$\begin{aligned} k_{AA}^e &= \int_0^h \left\{ \left[\psi_A^{e\prime}(\xi)\right]^2 + \left[\psi_A^e(\xi)\right]^2 \right\} d\xi \\ &= \int_0^h \left[\frac{1}{h^2} + \left(1 - \frac{\xi}{h}\right)^2 \right] d\xi = \frac{1}{h} + \frac{h}{3} \\ k_{AB}^e = k_{BA}^e &= \int_0^h \left\{ \psi_A^{e\prime}(\xi)\, \psi_B^{e\prime}(\xi) + \psi_A^e(\xi)\, \psi_B^e(\xi) \right\} d\xi \\ &= \int_0^h \left[\left(\frac{-1}{h}\right) \frac{1}{h} + \left(1 - \frac{\xi}{h}\right) \frac{\xi}{h} \right] d\xi = -\frac{1}{h} + \frac{h}{6} \end{aligned}$$

y

$$k_{BB}^{e}=\int_{0}^{h}\left\{\left[\psi_{B}^{e\prime}\left(\xi\right)\right]^{2}+\left[\psi_{B}^{e}\left(\xi\right)\right]^{2}\right\}d\xi=\frac{1}{h}+\frac{h}{3}$$

Similarmente, las componentes del vector de carga son

$$F_{A}^{e}=\int_{0}^{h}\left(x_{A}+\xi\right)\left(1-\frac{\xi}{h}\right)d\xi=\frac{h}{6}\left(2x_{A}+x_{B}\right)$$

y

$$F_{B}^{e}=\int_{0}^{h}\left(x_{A}+\xi\right)\left(\frac{\xi}{h}\right)d\xi=\frac{h}{6}\left(x_{A}+2x_{B}\right)$$

donde x_A y x_B son los valores de la función $f\left(x\right)=x$ en los nodos A y B. Estas cantidades son las entradas locales para la matriz de rigidez y del vector de carga $\mathbf{k}^e$ y $\mathbf{f}^e$ para el elemento Ω_e:

$$\mathbf{k}^{e}=\left[\begin{array}{cc}\frac{1}{h}+\frac{h}{3} & -\frac{1}{h}+\frac{h}{6}\\ -\frac{1}{h}+\frac{h}{6} & \frac{1}{h}+\frac{h}{3}\end{array}\right],\qquad\mathbf{f}^{e}=\frac{h}{6}\left[\begin{array}{c}2x_{A}+x_{B}\\ x_{A}+2x_{B}\end{array}\right]\tag{9.32}$$

Estos son los elementos de la matriz que de hecho pueden calcularse por un código informático de elementos finitos. Cuando la dimensión del problema esta especificada (en este ejemplo, habrá sólo tres ecuaciones) y cuando las coordenadas de los nodos en cada elemento sean especificadas, las entradas en (9.32) son calculadas y almacenadas en la fila i y la columna j apropiada para los nodos y los elementos que representan. De esta manera, la matriz de elementos ampliada $\mathbf{K}^e$ y $\mathbf{F}^e$ son entonces esencialmente por las sumas (9.30) y (9.31). Como $h=\frac{1}{4}$ en el presente ejemplo, el uso de (9.32) y el proceso que acabamos de describir proporciona las siguientes matrices de elementos ampliada:

$$\begin{aligned}\mathbf{K}^{1} &= \left[K_{ij}^{1}\right]=\frac{1}{24}\left[\begin{array}{ccc}98 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{array}\right],\qquad\mathbf{F}^{1}=\left\{F_{i}^{1}\right\}=\frac{1}{96}\left[\begin{array}{c}2\\ 0\\ 0\end{array}\right]\\ \mathbf{K}^{2} &= \left[K_{ij}^{2}\right]=\frac{1}{24}\left[\begin{array}{ccc}98 & -95 & 0\\ -95 & 98 & 0\\ 0 & 0 & 0\end{array}\right],\qquad\mathbf{F}^{2}=\left\{F_{i}^{2}\right\}=\frac{1}{96}\left[\begin{array}{c}4\\ 5\\ 0\end{array}\right]\\ \mathbf{K}^{3} &= \left[K_{ij}^{3}\right]=\frac{1}{24}\left[\begin{array}{ccc}0 & 0 & 0\\ 0 & 98 & -95\\ 0 & -95 & 98\end{array}\right],\qquad\mathbf{F}^{3}=\left\{F_{i}^{3}\right\}=\frac{1}{96}\left[\begin{array}{c}0\\ 7\\ 8\end{array}\right]\\ \mathbf{K}^{4} &= \left[K_{ij}^{4}\right]=\frac{1}{24}\left[\begin{array}{ccc}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 98\end{array}\right],\qquad\mathbf{F}^{4}=\left\{F_{i}^{4}\right\}=\frac{1}{96}\left[\begin{array}{c}0\\ 0\\ 10\end{array}\right]\end{aligned}$$

Por tanto, de acuerdo con (9.30) y (9.31),

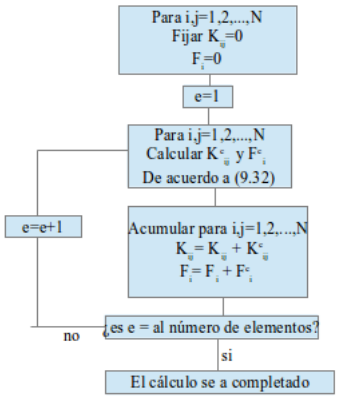
$$\mathbf{K} = [K_{ij}] = \mathbf{K}^1 + \mathbf{K}^2 + \mathbf{K}^3 + \mathbf{K}^4 = \frac{1}{24}\begin{bmatrix} 196 & -95 & 0 \\ -95 & 196 & -95 \\ 0 & -95 & 196 \end{bmatrix}$$

$$\mathbf{F} = \{F_i\} = \mathbf{F}^1 + \mathbf{F}^2 + \mathbf{F}^3 + \mathbf{F}^4 = \frac{1}{96}\begin{bmatrix} 6 \\ 12 \\ 18 \end{bmatrix}$$

y nuestro sistema final de ecuaciones es

$$\frac{1}{24}\begin{bmatrix} 196 & -95 & 0 \\ -95 & 196 & -95 \\ 0 & -95 & 196 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \frac{1}{16}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{9.33}$$

donde, de nuevo, u_1, u_2 y u_3 son los valores de u_h en los nodos $1, 2$ y 3 respectivamente. El procedimiento para obtener la matriz de rigidez en (9.33) puede ser descrito por el flujograma de la siguiente figura:



Una vez resuelto (9.33), encontramos que, con cuatro decimales

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0.0353 \\ 0.0569 \\ 0.0505 \end{bmatrix}$$

Así, por (9.25) la solución por elementos finitos a u_h de la ecuación (9.1) es

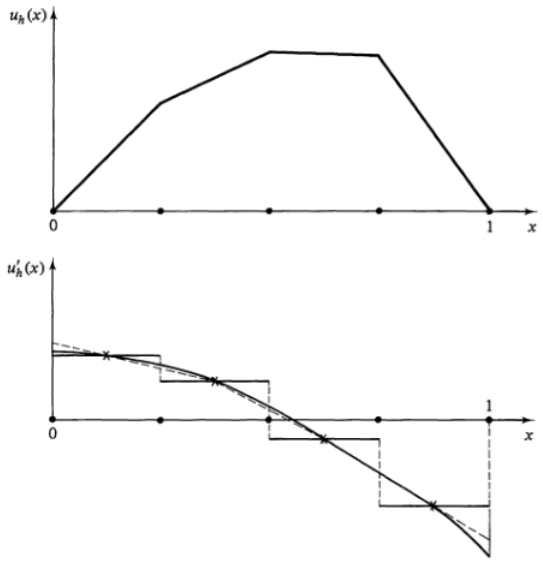
$$u_h(x) = 0.0353\phi_1(x) + 0.0569\phi_2 + 0.0505\phi_3(x)$$

donde las ϕ_i son las funciones mencionadas antes. Discutimos las propiedades de esta solución aproximada en la siguiente sección.

9.7. Interpretación de la solución aproximada

En las aplicaciones del método de elementos finitos, el análisis es particularmente complicado en el cálculo del vector de valores nodades **u**. Queda la tarea de interpretar la solución que ha sido encontrada. Las respuestas a preguntas cualitativas, tales como "¿cuál es el característica general de la solución?" o "¿dónde están las regiones en las que la solución varía más rápidamente?" son generalmente de interés, y estas se responden mejor examinando los gráficos de la solución y sus derivadas. El gráfico de la solución por elementos finitos no sólo muestran características cualitativas de la solución sino que también proporcionan una prueba fácil para la detección de errores de modelización de datos.

La aproximación por elementos finitos u_h de la solución del problema modelo descrita anteriormente y sus derivadas u_h' se muestran a continuación



Se hacen las siguientes observaciones:

1. La solución aproximada es una función bastante suave, no hay aparentes oscilaciones.
2. Hay un valor máximo 0.0569 localizado en $x = 0.5$.
3. La derivada de la solución es mayor cerca de los puntos finales, el mayor valor absoluto ocurre cerca de $x = 1.0$.

Por supuesto, si usamos una malla de elementos finitos "fina" (es decir, más elementos de menor longitud), entonces podríamos afinar nuestra imagen con estas características pero la aproximación que hemos calculado se adapta a nuestro propósito.

Para ver cómo estas características tienen relevancia en un problema físico, supongamos que el problema modelo ha surgido en el análisis de una cuerda, apoyada en una base elástica y sometida a una carga transversal cuya distribución está dada por $f(x) = x$. La solución u es la deflexión transversal de la cuerda y su derivada u' es proporcional a la tensión en la cuerda. Además, la energía de deformación total en el sistema (es decir, en la cuerda y en el soporte elástico) está dada por

$$U = \frac{1}{2}\int_0^1 \left[(u')^2 + u^2\right] dx. \tag{9.34}$$

En una aplicación así, podríamos buscar respuestas a las siguientes preguntas:

1. ¿cuál es la ubicación y el valor de máxima deflexión?
2. ¿cuál es la ubicación y el valor de máxima tensión?
3. ¿cuál es el valor de la energía de deformación total en el sistema?

La mejor respuesta a la pregunta 1 que podemos extraer de nuestra aproximación es que u_h tiene un pico en el nodo 2, y concluimos de esto que el valor de máxima deflexión ocurre en $x = 0.5$ y está dado por $u_2 = 0.0569$, como se señaló anteriormente. Esto es ligeramente un error, pero es la mejor información disponible en nuestra aproximación. La figura anterior indica que la respuesta a la pregunta 2 no es tan sencilla. Es claro que el valor máximo de $|u'|$ es 0.0202 y que este valor ocurre a lo largo del elemento Ω_4. Por motivos tanto físicos como matemáticos la tensión en el problema modelo varía continuamente. ¿A qué punto en el elemento Ω_4, entonces, vamos a asignar el valor calculado de estrés máximo? Una respuesta obvia es asignar el valor calculado para ser el valor medio del elemento y asignarlo al punto medio (es decir, al punto 0.875). Esta elección, es la mejor que podemos hacer, pero no llega a una respuesta satisfactoria para nuestra pregunta original.

La figura anterior muestra que deberíamos esperar que la tensión máxima ocurra en el punto $x = 1$. Con el fin de inferir, de la solución de elementos finitos, el valor de la derivada en el punto final $x = 1$, podríamos trazar, como en la figura el valor de la "tensión" u_h' en el punto medio del elemento. Podríamos extrapolar el valor de la tensión en el punto limite, $x = 1$, esbozando una curva suave a través de estos puntos por una curva continua, pero este procedimiento es demasiado impreciso para ser confiable y no es general. Alternativamente, podríamos ajustar una curva, digamos una linea recta por los puntos medios de u_h' para evaluar esta función en los puntos de interés . La linea punteada en la figura ilustra este proceso.

La energía de deformación de la solución por elementos finitos es fácilmente evaluada. La integral en (9.34) puede ser calculada elemento por elemento, como en la evaluación de la matriz de rigidez K_{ij}, y sumar los resultados sobre los elementos.

Sea ϕ el vector de funciones base y $\mathbf{u}$ el vector de valores nodales de la solución,

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Entonces $u_h(x) = \mathbf{u}^T \phi(x)$, donde $(\cdot)^T$ denota la transpuesta, y la energía aproximada es

$$\begin{aligned} U_h &= \frac{1}{2}\int_0^1 \left[(u_h')^2 + u_h^2\right] dx \\ &= \frac{1}{2}\int_0^1 \left[\phi'^T \phi' + \phi^T \phi\right] dx\, \mathbf{u} \\ &= \frac{1}{2}\mathbf{u}^T \mathbf{K}\mathbf{u}. \end{aligned}$$

Pero $\mathbf{K}\mathbf{u} = \mathbf{F}$, así

$$U_h = \frac{1}{2}\mathbf{u}^T \mathbf{F}. \tag{9.35}$$

Llevando a cabo los cálculos indicados en (9.35) dá el valor de la energía en la solución por elementos finitos de $U_h = 0.0094$.

Los cálculos y las observaciones tal como las indicadas muestran que una gran cantidad de información útil puede extraerse a partir de un examen cuidadoso de las propiedades de la solución.

10. Capítulo 7. Implementación de elementos finitos

En este capítulo estudiaremos el paper de Jochen Alberty, Carsten Carstensen y Stefan Funken, titulado "Remarks around 50 lines of Matlab: short finite element implementation".

Se proporciona una corta implementación en MATLAB para elementos finitos usando triángulos y paralelogramos para la solución numérica de problemas elípticos con condiciones de contorno mixtas en mallas no estructuradas. De acuerdo con la brevedad del programa y de la documentación facilitada, se puede realizar con facilidad cualquier adaptación de ejemplos modelo simples a problemas más complejos.

10.1. Problema modelo

El programa propuesto emplea el método de elementos finitos para calcular una solución numérica U que se aproxima a la solución u del problema bidimensional de Laplace (P) con condiciones de contorno mixtas: Sea $\Omega \subset \mathbb{R}^2$ con frontera poligonal Γ. En algún subconjunto cerrado Γ_D de la frontera, suponemos condiciones de Dirichlet, mientras que tenemos condiciones Neumann en la parte restante $\Gamma_N := \Gamma \setminus \Gamma_D$. Dada $f \in L^2(\Omega)$, $u_D \in H^1(\Omega)$ y $g \in L^2(\Gamma_N)$, buscar $u \in H^1(\Omega)$ con

$$-\triangle u = f \quad \text{en } \Omega \tag{10.1}$$

$$u = u_D \quad \text{en } \Gamma_D \tag{10.2}$$

$$\frac{\partial u}{\partial n} = g \quad \text{en } \Gamma_N. \tag{10.3}$$

De acuerdo con el lema de Lax Milgram, siempre existe una solución débil para (10.1) – (10.3). Las condiciones de Dirichlet no homogéneas (10.2) se incorporan a través de la descomposición $v = u - u_D$, así $v = 0$ en Γ_D, es decir,

$$v \in H^1_D(\Omega) := \left\{ w \in H^1(\Omega) \setminus w = 0 \text{ en } \Gamma_D \right\}.$$

Entonces, la formulación débil del problema de contorno (P) es: Buscar $v \in H_D(\Omega)$, de tal manera que

$$\int_\Omega \nabla v \cdot \nabla w \, dx = \int_\Omega f w \, dx + \int_{\Gamma_N} g w \, ds - \int_\Omega \nabla u_D \cdot \nabla w \, dx, \quad w \in H^1_D(\Omega). \tag{10.4}$$

10.2. Discretización de Galerkin del problema

Para la implementación, el problema (10.4) se discretiza utilizando el método de Galerkin, donde $H^1(\Omega)$ y $H^1_D(\Omega)$ se sustituyen por subespacios de dimensión finita S y $S_D = S \cap H^1_D$,

respectivamente.

Sea $U_D \in S$ una función que se aproxima u_D en Γ_D. (Se define U_D como la interpolante nodal de u_D en Γ_D.)
Entonces, el problema discretizado (P_S) es: Encontrar $V \in S_D$ tal que

$$\int_\Omega \nabla V \cdot \nabla W \, dx = \int_\Omega fW \, dx + \int_{\Gamma_N} gW \, ds - \int_\Omega \nabla U_D \cdot \nabla W \, dx, \quad W \in S_D. \tag{10.5}$$

Sea $(\eta_1, ..., \eta_N)$, una base del espacio de dimensión finita S, y sea $(\eta_{i_1}, ..., \eta_{i_M})$ una base de S_D, donde $I = \{i_1, ..., i_M\} \subseteq \{1, ..., N\}$ es un conjunto de índices de cardinalidad $M \leq N - 2$. Entonces, (10.5) es equivalente a

$$\int_\Omega \nabla V \cdot \nabla \eta_j \, dx = \int_\Omega f\eta_j \, dx + \int_{\Gamma_N} g\eta_j \, ds - \int_\Omega \nabla U_D \cdot \nabla \eta_j \, dx, \quad j \in I. \tag{10.6}$$

Además, sea

$$V = \sum_{k \in I} x_k \eta_k \qquad \text{y } U_D = \sum_{k=1}^{N} U_k \eta_k,$$

Entonces, de la ecuación (10.6) se obtiene el sistema lineal de ecuaciones

$$Ax = b. \tag{10.7}$$

La matriz de coeficientes $A = (A_{jk})_{j,k \in I} \in \mathbb{R}^{M \times M}$ y el vector del lado derecho $b = (b_j)_{j \in I} \in \mathbb{R}^M$ se definen como

$$\begin{aligned} A_{jk} &= \int_\Omega \nabla \eta_j \cdot \nabla \eta_k \, dx \\ b_j &= \int_\Omega f\eta_j \, dx + \int_{\Gamma_N} g\eta_j \, ds - \sum_{k=1}^{N} U_k \int_\Omega \nabla \eta_j \cdot \nabla \eta_k \, dx. \end{aligned} \tag{10.8}$$

La matriz de coeficientes es dispersa, simétrica y definida positiva, por lo que (10.7) tiene exactamente una solución $x \in \mathbb{R}^M$ que determina la solución Galerkin

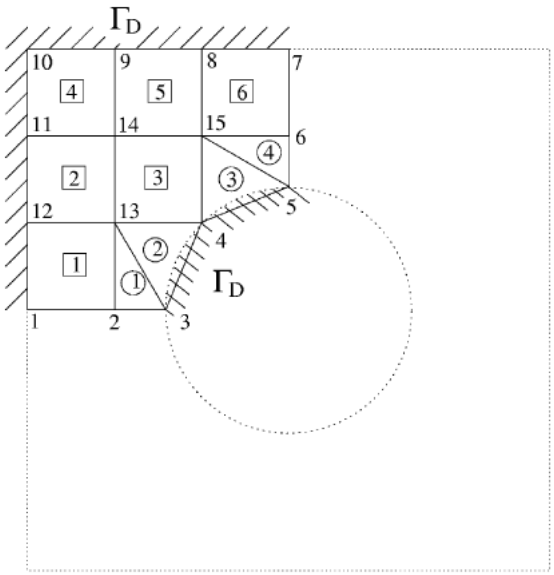
$$U = U_D + V = \sum_{j=1}^{N} U_j \eta_j + \sum_{k \in I} x_k \eta_k.$$

10.3. Representación de datos de la triangulación Ω.

Supongamos que el dominio Ω tiene una frontera poligonal Γ, podemos cubrir Ω por una triangulación regular $\mathcal{T}$ de triángulos y cuadriláteros, es decir, $\overline{\Omega} = \cup_{T \in \mathcal{T}} T$ y cada T es o bien un triángulo o un cuadrilátero.

La triangulación es regular, y esto significa que en los nodos $\mathcal{N}$ de la malla imaginaria en los vértices de los triángulos o cuadriláteros, los elementos de la triangulación no se superponen, ningún nodo se encuentra en un borde de un triángulo o cuadrilátero, y cada borde $E \subset \Gamma$ de un elemento $T \in \mathcal{T}$ pertenece ya sea a $\overline{\Gamma}_N$ o a $\overline{\Gamma}_D$.

MATLAB soporta la lectura de datos de archivos `.dat` en formato ascii.



La figura anterior muestra la malla que se describe por los siguientes datos. El archivo `coordinates.dat` contiene las coordenadas de cada nodo de la malla dada. Cada fila tiene la forma

nodo - coordenada x - coordenada y.

coordinates.dat

1	0	0
2	1	0
3	1.59	0
4	2	1
5	3	1.41
6	3	2
7	3	3
8	2	3
9	1	3
10	0	3
11	0	2
12	0	1
13	1	1
14	1	2
15	2	2

En nuestro código permitimos la subdivisión de Ω en triángulos y cuadriláteros. En ambos casos, los nodos están numerados en sentido antihorario. `elements3.dat` contiene para cada triángulo el número de nodo de los vértices. Cada fila tiene la forma

elemento - nodo1 - nodo2 - nodo3.

Del mismo modo, los datos de los cuadriláteros se dan en `elements4.dat`. Aquí, se utiliza el formato

elemento - nodo1 - nodo2 - nodo3 - nodo4

elements3.dat

1	2	3	13
2	3	4	13
3	4	5	15
4	5	6	15

elements4.dat

1	1	2	13	12
2	12	13	14	11
3	13	4	15	14
4	11	14	9	10
5	14	15	8	9
6	15	6	7	8

`neumann.dat` y `dirichlet.dat` contienen en cada fila los dos números de nodo que unen el borde correspondiente con la frontera:

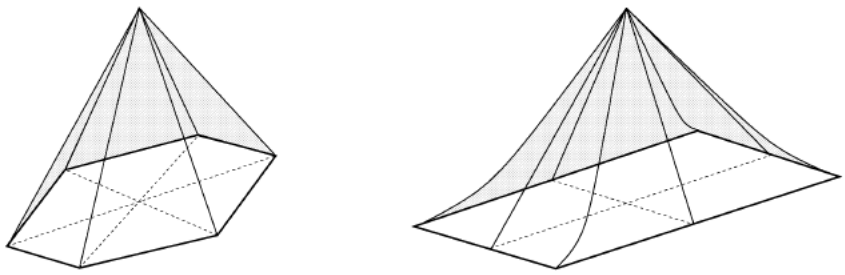
borde Neumann - nodo1- nodo2 y # borde Dirichlet - nodo1 - nodo2.

neumann.dat

1	5	6
2	6	7
3	1	2
4	2	3

dirichlet.dat

1	3	4
2	4	5
3	7	8
4	8	9
5	9	10
6	10	11
7	11	12
8	12	1



En la figura anterior mostramos dos funciones η_j que se definen para cada nodo (x_j, y_j) de la malla por

$$\eta_j\left(x_k, y_k\right) = \delta_{jk}, \quad j,k = 1, ..., N.$$

El subespacio $S_D \subset S$ es el espacio que es generado por todos aquellos η_j para los que (x_j, y_j) no estan en Γ_D. Entonces U_D, definida como la interpolante nodal de la u_D, se encuentra en S. Con estos espacios S, S_D y sus correspondientes bases, las integrales en (10.8) se puede calcular como una suma sobre todos los elementos y una suma sobre todos los bordes en Γ_D, es decir,

$$A_{jk} = \sum_{T\in\mathcal{T}} \int_T \nabla\eta_j \cdot \nabla\eta_k \, dx \tag{10.9}$$

$$b_j = \sum_{T\in\mathcal{T}} \int_T f\eta_j \, dx + \sum_{E\subset\Gamma_N} \int_E g\eta_j \, ds - \sum_{k=1}^{N} U_k \sum_{T\in\mathcal{T}} \int_T \nabla\eta_j \cdot \nabla\eta_k \, dx \tag{10.10}$$

10.4. Montaje de la matriz de rigidez

La matriz de rigidez local está determinada por las coordenadas de los vértices del elemento correspondiente y se calcula en las funciones `stima3.m` y `stima4.m`.

Para un elemento triangular T sean $(x_1, y_1), (x_2, y_2)$ y (x_3, y_3) los vértices y η_1, η_2 y η_3 las funciones de base correspondientes en S, es decir,

$$\eta_j(x_k, y_k) = \delta_{jk}, \qquad j, k = 1, 2, 3.$$

Que se puede escribir como:

$$\eta_j(x, y) = \det \begin{pmatrix} 1 & x & y \\ 1 & x_{j+1} & y_{j+1} \\ 1 & x_{j+2} & y_{j+2} \end{pmatrix} / \det \begin{pmatrix} 1 & x_j & y_j \\ 1 & x_{j+1} & y_{j+1} \\ 1 & x_{j+2} & y_{j+2} \end{pmatrix}, \tag{10.11}$$

por lo tanto

$$\nabla\eta_j(x, y) = \frac{1}{2|T|} \begin{pmatrix} y_{j+1} - y_{j+2} \\ x_{j+2} - x_{j+1} \end{pmatrix}.$$

Aquí, los índices han de entenderse en módulo 3, y $|T|$ es el área de T, es decir,

$$2|T| = \det \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix}.$$

La entrada resultante de la matriz de rigidez es

$$M_{jk} = \int_T \nabla\eta_j (\nabla\eta_k)^T dx = \frac{|T|}{(2|T|)^2} (y_{j+1} - y_{j+2}, x_{j+2} - x_{j+1}) \begin{pmatrix} y_{k+1} - y_{k+2} \\ x_{k+2} - x_{k+1} \end{pmatrix},$$

con índices en módulo 3. Esto se escribe simultáneamente para todos los índices como

$$M = \frac{|T|}{2} \cdot GG^T \quad \text{con } G = \begin{pmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

El siguiente procedimiento en MATLAB funciona simultáneamente para $d = 2$ y $d = 3$:

```
1   function M = stima3(vertices)
2 -   d = size(vertices,2);
3 -   G = [ones(1,d+1);vertices'] \ [zeros(1,d);eye(d)];
4 -   M = det([ones(1,d+1);vertices']) * G * G' / prod(1:d);
5
```

Para un elemento cuadrilátero T sean $(x_1, y_1), ..., (x_4, y_4)$ denotan los vértices con las correspondientes funciones $\eta_1, ..., \eta_4$. Dado que T es un paralelogramo,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \Phi_T(\xi, \zeta) = \begin{pmatrix} x_2 - x_1 & x_4 - x_1 \\ y_2 - y_1 & y_4 - y_1 \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} + \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

el cual es un mapeo de $[0, 1]^2$ sobre T. Entonces $\eta_j(x, y) = \varphi_j\left(\Phi_T^{-1}(x, y)\right)$ con funciones de forma

$$\begin{aligned}\varphi_1(\xi,\zeta) &:= (1-\xi)(1-\zeta), \qquad \varphi_2(\xi,\zeta) := \xi(1-\zeta),\\ \varphi_3(\xi,\zeta) &:= \xi\zeta, \qquad\qquad\quad \varphi_4(\xi,\zeta) := (1-\xi)\zeta,\end{aligned}$$

De la ley de sustitución se sigue para las integrales de (10.9) que

$$\begin{aligned}M_{jk}: &= \int_T \nabla\eta_j(x,y)\cdot\nabla\eta_k(x,y)\,dx(x,y)\\ &= \int_{(0,1)^2} \nabla\left(\varphi_j\circ\Phi_T^{-1}\right)(\Phi_T(\xi,\zeta))\left(\nabla\left(\varphi_k\circ\Phi_T^{-1}\right)\right)(\Phi_T(\xi,\zeta))^T\,|\det D\Phi_T|\,d(\xi,\zeta)\\ &= \det(D\Phi_T)\int_{(0,1)^2}\nabla\varphi_j(\xi,\zeta)\left((D\Phi_T)^T D\Phi_T\right)^{-1}(\nabla\varphi_k(\xi,\zeta))^T\,d(\xi,\zeta)\end{aligned}$$

Resolviendo estas integrales la matriz de rigidez local para un elemento cuadrilátero resulta en

$$M = \frac{\det(D\Phi_T)}{6}\begin{pmatrix} 3b+2(a+c) & -2a+c & -3b-(a+c) & a-2c\\ -2a+c & -3b+2(a+c) & a-2c & 3b-(a+c)\\ -3b-(a+c) & a-2c & 3b+2(a+c) & -2a+c\\ a-2c & 3b-(a+c) & -2a+c & -3b+2(a+c)\end{pmatrix}$$

donde

$$\left((D\Phi_T)^T(D\Phi_T)\right)^{-1} = \begin{pmatrix} a & b\\ b & c\end{pmatrix}.$$

```
1   function M = stima4(vertices)
2 -   D_Phi = [vertices(2,:)-vertices(1,:); vertices(4,:)- ...
3     vertices(1,:)]';
4 -   B = inv(D_Phi'*D_Phi);
5 -   C1 = [2,-2;-2,2]*B(1,1)+[3,0;0,-3]*B(1,2)+[2,1;1,2]*B(2,2);
6 -   C2 = [-1,1;1,-1]*B(1,1)+[-3,0;0,3]*B(1,2)+[-1,-2;-2,-1]*B(2,2);
7 -   M = det(D_Phi) * [C1 C2; C2 C1] / 6;
8
```

10.5. Montaje de la parte derecha

Las fuerzas de volumen se utilizan para el montaje del lado derecho. Usando el valor de f en el centro de gravedad (x_S, y_S) de T la integral $\int_T f\eta_j\,dx$ en (10.10) se aproxima por

$$\int_T f\eta_j\,dx = \frac{1}{k_T}\det\begin{pmatrix} x_2-x_1 & x_3-x_1\\ y_2-y_1 & y_3-y_1\end{pmatrix} f(x_S,y_S)$$

donde $k_T = 6$ si T es un triángulo y $k_T = 4$ si T es un paralelogramo.

```
% Volume Forces
for j = 1:size(elements3,1)
    b(elements3(j,:)) = b(elements3(j,:)) + ...
    det([1 1 1; coordinates(elements3(j,:),:)']) * ...
    f(sum(coordinates(elements3(j,:),:))/3)/6;
end

for j = 1:size(elements4,1)
    b(elements4(j,:)) = b(elements4(j,:)) + ...
    det([1 1 1; coordinates(elements4(j,1:3),:)'])* ...
    f(sum(coordinates(elements4(j,:),:))/4)/4;

end
```

Los valores de f son dados por la función `f.m` que depende del problema.

La función se llama con las coordenadas de los puntos en Ω y devuelve las fuerzas de volumen en estos lugares. Para el ejemplo numérico mostrado en la figura anterior se utilizó

```
function VolumeForce = f(x)
VolumeForce = ones(size(x,1),1);

```

Del mismo modo, las condiciones Neumann contribuyen al lado derecha. La integral $\int_E g\eta_j ds$ en (10.10) se aproxima utilizando el valor de g en el centro (x_M, y_M) de E con una longitud $|E|$ por

$$\int_E g\eta_j ds \approx \frac{|E|}{2} g(x_M, y_M)$$

```
% Neumann conditions
for j = 1 : size(neumann,1)
    b(neumann(j,:))=b(neumann(j,:)) + ...
    norm(coordinates(neumann(j,1),:) - ...
    coordinates(neumann(j,2),:)) * ...
    g(sum(coordinates(neumann(j,:),:))/2)/2;
end
```

Aquí se utiliza el hecho de que en MATLAB el tamaño de una matriz vacía se fija igual a cero y que un bucle de 1 a 0 se omite totalmente. De esa manera, la existencia de condiciones Neumann no se pone.

Los valores de g están dados por la función `g.m` que depende de nuevo en el problema. La función se llama con las coordenadas de los puntos en Γ_N y devuelve las tensiones correspondientes. Para el ejemplo numérico `g.m` era

```
function Stress = g(x)
Stress = zeros(size(x,1),1);

```

10.6. Incorporación de las condiciones Dirichlet

Con una adecuada numeración de los nodos, el sistema de ecuaciones lineales que resulta de la construcción descrita en la sección anterior sin incorporar condiciones de Dirichlet se puede escribir como sigue:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \cdot \begin{pmatrix} U \\ U_D \end{pmatrix} = \begin{pmatrix} b \\ b_D \end{pmatrix}, \tag{10.12}$$

con $U \in \mathbb{R}^M$, $U_D \in \mathbb{R}^{N-M}$. Aquí, U son los valores en los nodos libres los cuales seran determinados, U_D son los valores en los nodos que están en la frontera Dirichlet y por lo tanto se conocen a priori. Por lo tanto, el primer bloque de ecuaciones puede reescribirse como

$$A_{11} \cdot U = b - A_{12} \cdot U_D.$$

De hecho, esta es la formulación de (10.6) con $U_D = 0$ en los nodos que no son Dirichlet.

En el segundo bloque de ecuaciones en (10.12) lo desconocido es b_D pero ya que no es de interés para nosotros se omite en lo sucesivo.

```
1     % Dirichlet conditions
2 -   u = sparse(size(coordinates,1),1);
3 -   u(unique(dirichlet)) = u_d(coordinates(unique(dirichlet),:));
4 -   b = b - A * u;
5
```

Los valores u_D de los nodos en Γ_D están dados por la función `u_d.m` que depende del problema. La función se llama con las coordenadas de los puntos en Γ_D y devuelve los valores en los lugares correspondientes. Para el ejemplo numérico `u_d.m` fue

```
1     function DirichletBoundaryValue = u_d(x)
2 -   DirichletBoundaryValue = zeros(size(x,1),1);
3
```

10.7. Cálculo y visualización de la solución numérica

Las filas de (10.7) correspondientes a las primeras M filas de (10.12) forman un sistema de ecuaciones reducido con una matriz de coeficientes simétrica y definida positiva M_{11}. Esta se obtiene a partir del sistema original de ecuaciones tomando las filas y las columnas correspondientes a los nodos libres del problema. La restricción se puede lograr en MATLAB a través de la indexación adecuada.

El sistema de ecuaciones se resuelve en MATLAB con el operador binario \ que da la inversa de una matriz.

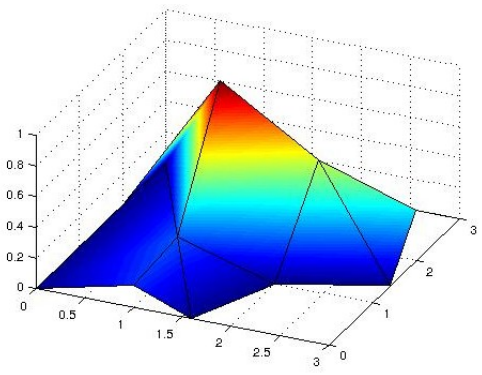
```
1 -   FreeNodes=setdiff(1:size(coordinates,1),unique(dirichlet));
2 -   u(FreeNodes)=A(FreeNodes,FreeNodes)\b(FreeNodes);
3
```

MATLAB hace uso de las propiedades de una matriz simétrica, definida positiva y dispersa para resolver el sistema de ecuaciones de manera eficiente.

Una representación gráfica de la solución puede darse por la función `show.m`.

```
1   function show(elements3,elements4,coordinates,u)
2 -   trisurf(elements3,coordinates(:,1),coordinates(:,2),u',...
3         'facecolor','interp')
4 -   hold on
5 -   trisurf(elements4,coordinates(:,1),coordinates(:,2),u',...
6         'facecolor','interp')
7 -   hold off
8 -   view(10,40);
9 -   title('Solution of the Problem')
```

Aquí, el procedimiento de MATLAB `trisurf(elementos,X,Y,U)` se utiliza para dibujar triangulaciones para tipos iguales de elementos. Cada fila de la matriz `elements` determina un polígono en el que las coordenadas x, y, y z de cada esquina de este polígono está dada por la entrada correspondiente en X, Y y U, respectivamente. El color de los polígonos está dado por los valores de U. Los parámetros adicionales, 'facecolor', 'interp', conducen a una coloración interpolada.



La figura muestra la solución para la malla definida al principio y los archivos de datos `f.m`, `g.m`, y `u_d.m` definidos como antes.
Resumiendo lo anterior, el programa principal, se estructura de la siguiente manera:

- Líneas 1-9: Carga de la geometría de la malla y la inicialización.
- Líneas 10-18: Ensamble de la matriz de rigidez, primero sobre los elementos triangulares y luego en los cuadriláteros.
- Líneas 19-27: La incorporación de la fuerza de volumen, primero sobre elementos triangulares y luego en los cuadriláteros.
- Líneas 29-32: La incorporación de la condición Neumann.
- Líneas 33-35: La incorporación de la condición Dirichlet.
- Líneas 36-37: La solución del sistema lineal.
- Líneas 38-39: Representación gráfica de la solución numérica.

10.8. Código

El código para el problema de Laplace de 2 dimensiones

El programa siguiente se llama `fem2d.m`. Los otros archivos bajo ese camino son las funciones `stima3.m, stima4.m` y `show.m`, así como las funciones y los archivos de datos que describen la discretización y los datos del problema, a saber `coordinates.dat, elements3.dat, elements4.dat, dirichlet.dat, neumann.dat, f.m, g.m` y `u_d.m`. Esos archivos describiendo el problema deben ser adaptados por el usuario para otras geometrías, discretizaciones, y / o datos.

```
% Initialisation
load ('coordinates.dat'); coordinates(:,1)=[];
eval('load elements3.dat; elements3(:,1)=[];','elements3=[];');
eval('load elements4.dat; elements4(:,1)=[];','elements4=[];');
eval('load neumann.dat; neumann(:,1) = [];','neumann=[];');
load dirichlet.dat; dirichlet(:,1) = [];
FreeNodes=setdiff(1:size(coordinates,1),unique(dirichlet));
A = sparse(size(coordinates,1),size(coordinates,1));
b = sparse(size(coordinates,1),1);
% Assembly
for j = 1:size(elements3,1)
  A(elements3(j,:),elements3(j,:)) = A(elements3(j,:),elements3(j,:)) ...
      + stima3(coordinates(elements3(j,:),:));
end
for j = 1:size(elements4,1)
  A(elements4(j,:),elements4(j,:)) = A(elements4(j,:),elements4(j,:)) ...
      + stima4(coordinates(elements4(j,:),:));
end
% Volume Forces
for j = 1:size(elements3,1)
  b(elements3(j,:)) = b(elements3(j,:)) + det([1,1,1; coordinates(elements3(j,:),:)']) * ...
      f(sum(coordinates(elements3(j,:),:))/3)/6;
end
for j = 1:size(elements4,1)
  b(elements4(j,:)) = b(elements4(j,:)) + det([1,1,1; coordinates(elements4(j,1:3),:)']) * ...
      f(sum(coordinates(elements4(j,:),:))/4)/4;
end
% Conditions
for j = 1 : size(neumann,1)
  b(neumann(j,:))=b(neumann(j,:)) + norm(coordinates(neumann(j,1),:)- ...
      coordinates(neumann(j,2),:)) * g(sum(coordinates(neumann(j,:),:))/2)/2;
end
u = sparse(size(coordinates,1),1);
u(unique(dirichlet)) = u_d(coordinates(unique(dirichlet),:));
b = b - A * u;
% Computation of the solution
u(FreeNodes) = A(FreeNodes,FreeNodes) \ b(FreeNodes);
% graphic representation
show(elements3,elements4,coordinates,full(u));
```

11. Conclusiones y Resultados

Debido a las múltiples aplicaciones que tienen las ecuaciones diferenciales, es de fundamental importacia estudiar, analizar e implementar métodos y algoritmos para resolver diferentes problemas de éstas de manera numérica y obtener resultados satisfactorios.

Estos programas y códigos computacionales, se hacen basados en dos métodos que son eficientes, como el método de diferencias finitas y el método de elementos finitos, el primero se lleva a cabo sustituyendo la derivada en la ecuación diferencial por una aproximación en diferencias de la derivada y se reescribe la ecuación en forma que se pueda apreciar que valores queremos calcular en términos de los que ya conocemos. Este método puede llevar a esquemas que pueden ser explícitos, en donde se calcula la aproximación de manera directa, o implícitos cuando se nos forma un sistema de ecuaciones que resolvemos por métodos conocidos. El método de elementos finitos se lleva a cabo buscando la solución en un espacio adecuado y reescribiendo el problema en su variacional equivalente y al final se reduce a un sistema lineal de ecuaciones.

En la teoría general de solución numérica de ecuaciones diferenciales, no solo interesa aproximar de manera numérica la solución de una ecuación diferencial, por que una aproximación sólo es útil si es convergente, es decir a medida que se decrese el tamaño de paso nos cercamos a la solución exacta. Como un requisito para la convergencia tenemos la consistencia. En el método de elementos finitos, al refinar la malla (elementos más pequeños), la solución tiende hacia la solución exacta y de esa forma se muestra la convergencia.

La implementación de ambos métodos se hizó en Octave y todos los códigos mostrados son compatibles con MATLAB, por ser lenguajes de alto desempeño diseñados para realizar cálculos técnicos, que integran el cálculo, la visualización y la programación en un ambiente fácil de utilizar. A la vez los códigos pueden adaptarse a otra ecuación diferencial en el código proporcionado.

12. Referencias

[1] Douglas N. Arnold, A Consice Introduction to Numerical Analysis. School of Mathematics, University of Minnesota, Minneapolis.

[2] Cheney, Ward y Kincaid, David. Numerical Mathematics ans Computing. The University of Texas at Austin.

[3] Burden, Richard y Faires, Douglas. Analisis Numérico. Thomson learning, 2002.

[4] E. Gorach, E. Andrunov. Notas de Analisis Funcional. UBA

[5] R. Bruzual, M. Dominguez. Espacios de Hilbert. Universidad Central de Venezuela

[6] H. Falomir. Curso de métodos de la Física Matemática. Facultad de Ciencias Exactas, UNLP.

[7] Kendall Atkinson, Weimin Han. Theoretical Numerical Analysis, Third edition, Springer

[8] Jasper Schmidt Hansen, GNU Octave beginners guide.

[9] Mark Goctenbach, MATLAB Tutorial.

[10] Schaerer, Christian, Introducción a los métodos numéricos para EDP, Universidad Nacional de Asunción, Paraguay.

[11] C. Johnson. Numerical Solution of Partial Differential Equation by the Finite Element Method, Cambridge University Press, 1987.

[12] J.T Oden, E.B. Becker y G.F Caley. Finite Elements: An Introduction. Volume I. Prentice Hall, 1981.

[13] Rodolfo Rodriguez. Notas del curso Solución Numérica por Elementos Finitos de Ecuaciones Diferenciales Parciales, impartido en EMALCA Turrialba, 2014.

[14] J. Alberty, C. Carstensen y S.A. Funken. Remarks around 50 Lines of Matlab: short finite element implementation. Numerical Algorithms, (1999).

[15] M. Maleco, A. Salvador, T. Menarguez, L. Garmendia. Análisis Matemático para Ingeniería. Pág. 368.

[16] John C. Strkwerda. Finite difference schemes and partial differential equations. Second Edition.