

# Aspectos físicos y matemáticos del método del elemento finito

Gerardo Aragón

*PDPA, Universidad Autónoma Metropolitana-Azcapotzalco  
San Pablo, 180, La Reynosa Azcapotzalco, 02200 México, D.F., Mexico*

José Luis Aragón

*Instituto de Física, Universidad Nacional Autónoma de México  
Apartado postal 20-364, 01000 México, D.F., Mexico*

Recibido el 3 de marzo de 1997; aceptado el 9 de febrero de 1998

Por lo general, el método de elemento finito se presenta destacando solamente ya sea aspectos físicos o matemáticos. El propósito de este trabajo es mostrar algunos aspectos físicos y matemáticos de este método en una forma integrada. El trabajo puede servir como una introducción a temas más avanzados y específicos de este método.

*Descriptores:* Métodos de elemento finito y Galerkin; técnicas computacionales; educación

In most of the introductory works on the finite element method, either physical aspects or mathematical foundations are stressed. The main objective of this work is to present a unified view of this method which considers both physical and mathematical points of view. This work can also serve as an introduction to more advanced topics.

*Keywords:* Finite-element and Galerkin methods; computational techniques; education

PACS: 02.70.Dh; 02.70.-c; 01.40.-d

## 1. Introducción

La mayoría de los fenómenos físicos, ya sean biológicos, geológicos o mecánicos, pueden ser descritos con ayuda de leyes físicas que se traducen en ecuaciones diferenciales, integrales o algebraicas. Aun cuando la deducción de estas ecuaciones no sea difícil, encontrar soluciones, vía métodos exactos, presenta diversas complejidades. En estos casos los métodos de aproximación representan la mejor alternativa para encontrar soluciones aproximadas. Entre los métodos más comúnmente usados se encuentra el *método de diferencias finitas* y los *métodos variacionales* como el de Ritz y Galerkin [1]. Recientemente ha cobrado relevancia el denominado *método de elemento finito*.

La aproximación vía diferencias finitas de una ecuación diferencial consiste en reemplazar las derivadas por cocientes en diferencias, los cuales consideran los valores de la solución en puntos que corresponden a una malla discreta del dominio. Las ecuaciones algebraicas que resultan se resuelven, contemplando las condiciones de frontera, para valores de la solución en puntos de la malla. Aun cuando esta técnica es muy útil y ha sido muy empleada para resolver ciertos problemas, presenta dificultades cuando las geometrías son irregulares o cuando cambian las condiciones de frontera.

Cuando es posible transformar las ecuaciones diferenciales a una forma variacional o débil, la solución a este problema variacional puede hallarse en una forma aproximada al suponer que la solución puede escribirse como una combinación  $\sum \alpha_i \phi_i$  de funciones de aproximación  $\phi_i$  que han sido elegidas de antemano. Los parámetros  $\alpha_i$  se calculan

con base en la sustitución de esta combinación en la forma variacional, que en algunas ocasiones queda sujeta a un proceso de minimización. Como resultado de este proceso surge un sistema de ecuaciones en donde las incógnitas son las  $\alpha_i$ . Sin embargo, si los dominios son arbitrarios las funciones de aproximación son difíciles de elegir o construir debido a que los métodos empleados (Ritz y Galerkin) no proporcionan una manera sistemática para construir dichas funciones.

El método de elemento finito proporciona un procedimiento sistemático para encontrar funciones de aproximación que puede ser aplicado a dominios con geometría irregular. En este método se puede representar un dominio complejo, en donde el fenómeno que nos interesa está ocurriendo como una colección de subdominios (elementos finitos). Las funciones de aproximación se pueden construir en cada uno de estos elementos finitos (se pueden considerar, incluso, elementos diferentes, con diferente forma de la función de aproximación). La idea esencial que está detrás de este proceso es que cualquier función continua sobre el elemento puede ser representada por una combinación lineal (interpolación) de polinomios; de hecho, las funciones de aproximación se encuentran usando conceptos de la teoría de interpolación.

Con el método de elemento finito, entonces, estamos estudiando una región del espacio, en donde el fenómeno de interés está ocurriendo, con base en sus elementos finitos. El análisis se realiza sobre solamente uno de los elementos (en el caso de que sólo haya un tipo de elemento) cuya geometría es sencilla. Después se procede a coleccionar a todos los elementos involucrados y con ello encontramos una solución aproximada al problema.

En este trabajo mostraremos las ideas principales del método del elemento finito y nuestra discusión iniciará con el análisis de un problema antiguo: encontrar el área del círculo mediante sucesiones de polígonos regulares inscritos y circunscritos al círculo. Posteriormente, estudiaremos el fenómeno de conducción de calor en una barra para mostrar lo que se conoce como *aproximación directa o física* [2] y, si incluimos en el problema la generación interna de calor, la *aproximación mediante balance energético* [3]. Por último, presentamos algunos aspectos matemáticos adheridos al método de elemento finito, enfocados en la aproximación variacional-residuos pesados, apoyados en el método de Galerkin [1, 4]. La presentación pretende ser didáctica y con este fin el análisis se hará solamente en una dimensión, que es suficiente para describir las técnicas que subyacen en el método de elemento finito.

## 2. Un problema antiguo y el método del elemento finito

No aparece por primera vez con el método de elemento finito la idea de representar un dominio dado como una colección de elementos discretos. Esta idea se remonta a la antigüedad con las matemáticas griegas. Arquímedes (siglo I D.C.) hace una estimación del valor de  $\pi$  hasta 40 dígitos, al darse cuenta que el perímetro de un círculo se podía aproximar mediante el perímetro de polígonos regulares; uno inscrito y otro circunscrito. Arquímedes pudo así representar al círculo como una sucesión de longitudes de polígonos regulares con un número creciente de lados. El perímetro de la circunferencia será el límite de esa sucesión de longitudes y se denota por  $2\pi$ . Específicamente, si  $p_n$  representa la longitud del polígono inscrito de  $n$  lados y  $q_n$  el correspondiente circunscrito, entonces

$$p_n < 2\pi < q_n,$$

aquí como  $n$  aumenta,  $p_n$  y  $q_n$  se aproximan monóticamente a  $2\pi$ . A partir de la expresión

$$p_{2^m} = 2^m \sqrt{2 + \sqrt{2 + \sqrt{2 + \dots + \sqrt{2}}}},$$

podemos calcular aproximadamente el valor de  $2\pi$  [7].

Determinemos el área de un círculo de radio  $R$  mediante aproximación de esta área por medio de polígonos regulares inscritos y circunscritos. Podemos considerar el círculo como colecciones de triángulos (Fig. 1). En este caso, la suma de las áreas de los triángulos empleados para representar el círculo será el área aproximada de este círculo. De esta manera, el círculo queda representado con dos colecciones o mallas uniformes de elementos (pues los elementos en cada malla son idénticos) que constan de un número finito de triángulos: los inscritos  $T^{(e)}$  y los circunscritos  $\bar{T}^{(e)}$ . Consideremos dos elementos típicos, uno de cada malla,  $T^{(e)}$  y  $\bar{T}^{(e)}$ , los cuales consideraremos aislados y calcularemos sus propiedades, es decir, sus áreas (Fig. 2). Si  $a^{(e)}$  representa el área del triángulo  $T^{(e)}$  y  $\bar{a}^{(e)}$  el área de  $\bar{T}^{(e)}$ .

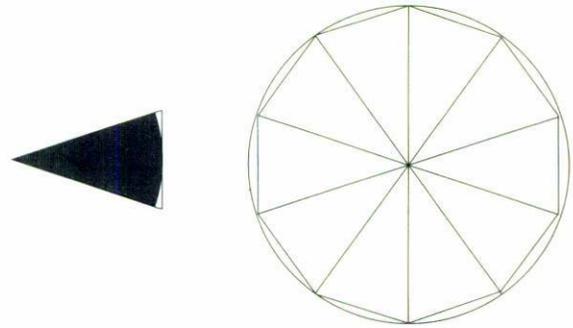


FIGURA 1. Aproximación del círculo mediante polígonos regulares.

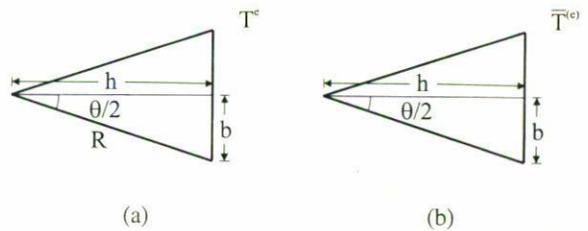


FIGURA 2. Elemento finito para el cálculo del área del círculo para ambos elementos.

Para cada elemento ( $e$ ), encontramos:

$$\begin{aligned} \text{para } T^{(e)}: & \quad b = R \sin \frac{\theta}{2} & \quad h = R \cos \frac{\theta}{2} \\ \text{para } \bar{T}^{(e)}: & \quad h = R & \quad b = R \tan \frac{\theta}{2}. \end{aligned}$$

Como son polígonos regulares de  $n$  lados, el ángulo  $\theta$  está dado por  $\theta = 2\pi/n$ , así el área de los triángulos es

$$a^{(e)} = \frac{R^2}{2} \sin \frac{2\pi}{n} \quad \text{y} \quad \bar{a}^{(e)} = R^2 \tan \frac{\pi}{n}.$$

Ahora sumemos las áreas de los elementos individuales

$$A_n^{(1)} = \sum_{e=1}^n a^{(e)} \quad \text{y} \quad A_n^{(2)} = \sum_{e=1}^n \bar{a}^{(e)}, \quad (1)$$

que son iguales a

$$A_n^{(1)} = n \frac{R^2}{2} \sin \frac{2\pi}{n} \quad \text{y} \quad A_n^{(2)} = n R^2 \tan \frac{\pi}{n}.$$

La ventaja de este problema es que conocemos la solución exacta, que es  $A^{(0)} = \pi R^2$ , por lo que podemos estimar el error en esta aproximación. Los errores de la aproximación son iguales a la diferencia entre el área del sector circular  $S^{(e)}$  y la correspondiente del triángulo (Fig. 3):  $|S^{(e)} - a^{(e)}|$  y  $|S^{(e)} - \bar{a}^{(e)}|$ , con  $S^{(e)} = \frac{1}{2} R^2 \theta$  (área del sector). Así,

$$\begin{aligned} |S^{(e)} - a^{(e)}| &= R^2 \left( \frac{\pi}{n} - \frac{1}{2} \sin \frac{2\pi}{n} \right), \\ |S^{(e)} - \bar{a}^{(e)}| &= R^2 \left( \tan \frac{\pi}{n} - \frac{\pi}{n} \right). \end{aligned}$$

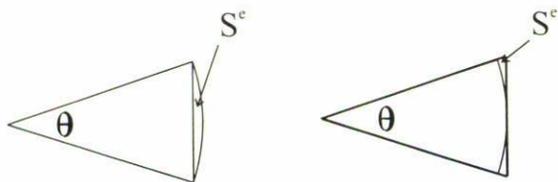


FIGURA 3. Exceso y defecto en la aproximación del círculo por triángulos.

Los errores globales se obtienen multiplicando por  $n$  estas expresiones:

$$R^2 \left( \pi - \frac{n}{2} \operatorname{sen} \frac{2\pi}{n} \right) = \pi R^2 - A_n^{(1)},$$

$$R^2 \left( n \tan \frac{\pi}{n} - \pi \right) = A_n^{(2)} - \pi R^2,$$

las cuales tienden a cero cuando  $n \rightarrow \infty$ . En efecto, sean  $x = 2/n$  y  $y = 1/n$ , entonces

$$\lim_{n \rightarrow \infty} A_n^{(1)} = \lim_{x \rightarrow 0} R^2 \frac{\operatorname{sen} \pi x}{x} = \pi R^2,$$

$$\lim_{n \rightarrow \infty} A_n^{(2)} = \lim_{y \rightarrow 0} R^2 \frac{\tan \pi y}{y} = \pi R^2.$$

Conviene hacer un resumen del proceso anterior. Los pasos más importantes que hemos seguido pueden resumirse como sigue: la región o dominio continuo (círculo) lo hemos representado por medio de una colección de un número finito  $n$  de subdominios (triángulos). Hemos logrado así una *discretización del dominio*. A cada una de las subregiones las llamaremos *elementos finitos*, y a la colección *mallado de elementos finitos* (en el ejemplo hemos propuesto dos mallas, y como los elementos son iguales, ésta se dice *uniforme*).

Podemos discretizar el dominio en distintas mallas con diferentes tipos de elementos que dependerán de la forma geométrica del dominio. En la aproximación del área del círculo se pueden utilizar combinaciones de rectángulos y triángulos para aproximarnos “más cercanamente”.

Después, un elemento típico (triángulo o rectángulo) es aislado, y sus propiedades (áreas) son desarrolladas. Podemos llamarlas *ecuaciones de los elementos* (ecuaciones que sirven para calcular el área). Si consideramos más de un tipo de elemento en la representación del dominio, hay que aislar cada clase y desarrollar las propiedades para cada uno de los elementos típicos.

Obtuvimos el área aproximada del círculo al “reunir” las propiedades de los elementos finitos; este proceso se conoce como *ensamblaje*. En este ejemplo, el ensamblaje consistió en sumar las áreas de los elementos individuales [Ec. (1)]. Como en este problema conocemos la solución exacta, estimar el error de la aproximación no presentó ninguna dificultad, sin embargo, habitualmente no conocemos la solución exacta y estimar el error en estos casos es una tarea compleja desde el punto de vista matemático. Además, sólo hemos considerado errores que provienen de aproximar el dominio; en general hay que considerar otros tipos de errores.

Con este ejemplo sencillo hemos ilustrado algunas de las ideas principales del método de elemento finito. Falta, sin embargo, explicar otros hechos que son importantes y que aún no hemos considerado. Busquemos expresar estos hechos a través de la discusión de la conducción de calor en una barra.

### 3. Elemento finito y conducción de calor

La conducción térmica es un fenómeno de transporte en el que se transfiere energía de una región a otra con un cambio correspondiente de la temperatura. La conducción tiene su origen en una diferencia de temperatura, la cual produce un flujo neto de energía en el sentido en que disminuye la temperatura.

La conducción térmica es un proceso meramente molecular, por lo que ocurre en forma diferente en sólidos, líquidos y gases debido a la diferencia de movilidad molecular en los tres estados. En los gases, y en cierta medida en los líquidos, la conducción térmica es el resultado de colisiones entre las moléculas. En regiones en que la temperatura es más alta, las moléculas tienen velocidades mayores que en regiones con temperatura menor. Existen, entonces, colisiones entre moléculas que se mueven más rápido y las que se mueven más lento, lo que produce una transferencia de energía cinética de las moléculas más rápidas hacia las más lentas. La energía transferida a la región de moléculas con movimiento más lento (las cuales se encuentran en la región con temperatura menor) se denomina *calor*.

En cambio en los sólidos la situación es distinta, al no existir movimiento molecular, el único movimiento se produce por la vibración de las moléculas alrededor de sus posiciones de equilibrio. En regiones con mayor temperatura, la energía vibracional es mayor que en regiones con temperatura menor y, como consecuencia, es esta energía vibracional la que se transporta a lo largo de todo el sólido. En los metales existe un efecto adicional, ya que los electrones de conducción pueden moverse libremente en todo el volumen del metal. Estos electrones se comportan en forma similar a las moléculas en un gas y tienden a difundirse a través del metal de la región caliente a la fría; transfiriendo energía por colisión con otros electrones y con los iones de la red en la región más fría.

Desde el punto de vista macroscópico, la conducción térmica o conducción de calor depende básicamente de las propiedades físicas del cuerpo, su forma geométrica y la diferencia entre las temperaturas de las diferentes partes del cuerpo.

La relación básica para la transferencia de calor por conducción de calor fue propuesta en 1822 por el científico francés Joseph Fourier (1768-1830). En ese entonces el flujo de calor era de interés práctico para el manejo de metales y de interés científico porque se quería determinar la temperatura en el interior de la Tierra.

Para analizar el fenómeno de conducción de calor consideraremos una pared homogénea, cuyos extremos se mantienen a temperaturas constantes  $T_1$  y  $T_2$  ( $T_1 > T_2$ ). Supondre-

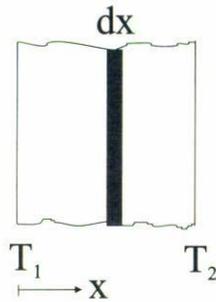


FIGURA 4. Conducción unidimensional de calor en una pared homogénea sin generación interna de calor.

mos que la temperatura sólo varía en la dirección  $x$  y que no hay generación interna de calor. Las superficies isotermas son las secciones transversales al eje  $x$  (Fig. 4). A una distancia  $x$ , en el interior de la pared, separemos una capa de espesor  $dx$ . Llamemos  $q$  a la densidad de corriente de energía (densidad de flujo de calor) debida a la diferencia de temperatura, es decir, la energía que atraviesa en la unidad de tiempo un área unitaria perpendicular a la dirección en que tiene lugar el flujo de energía. Según vimos arriba, este flujo de energía tiene lugar en el sentido en el que la temperatura disminuye. El cambio de temperatura por unidad de longitud (o gradiente de temperatura) del material es, entonces,  $dT/dx$ . La ley de Fourier establece que, a menos que la temperatura varíe muy rápidamente en una distancia corta,  $q$  es proporcional a  $dT/dx$ , es decir,

$$q = -k \frac{dT}{dx},$$

donde  $k$ , la conductividad térmica del material, expresa las propiedades físicas de éste y caracteriza su capacidad para conducir calor. El signo negativo indica que la energía fluye en el sentido en que la temperatura disminuye. Aunque el mecanismo de conducción de calor es diferente en gases, líquidos y sólidos, la ley de Fourier se aplica a estos tres estados de la materia.

Si consideramos a  $k$  como constante [generalmente  $k = k(x)$ ], de la ley de Fourier

$$dT = -\frac{q}{k} dx,$$

pero si es régimen estacionario,  $q$  es constante en cada sección, entonces

$$T = -\frac{q}{k} x + \text{cte.}$$

La constante la podemos determinar a partir de las condiciones de frontera:  $x = 0$ ,  $T = T_1$ ,  $x = L$  y  $T = T_2$  ( $L$  es la longitud de la barra). Así,

$$T_2 = -\frac{q}{k} L + T_1,$$

o bien

$$q = \frac{k}{L} (T_1 - T_2) = \frac{k}{L} \Delta T.$$

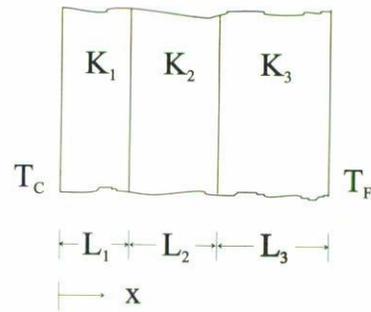


FIGURA 5. Discretización de la pared en donde se realiza la conducción de calor.

“La cantidad de calor transmitida a través de una unidad de superficie es directamente proporcional al coeficiente de conductividad térmica y a la diferencia de temperatura  $\Delta T$  que existe en los extremos de la pared, e inversamente proporcional al espesor de la pared”, por lo que

$$T(x) = T_1 - \left( \frac{T_1 - T_2}{L} \right) x.$$

La temperatura en una pared homogénea varía linealmente si el coeficiente de conductividad térmica es constante.

Aunque esto último no siempre sucede, pues existe una fuerte dependencia de la conductividad térmica con la temperatura y, por ende, con  $x$ . Aun así, podemos aplicarlo a paredes multilaminares: muros formados por varias capas heterogéneas, denominados muros de capas múltiples, por ejemplo, paredes de edificios, revestimientos de hornos y calderas, etc. Para analizar este problema emplearemos algunas de las ideas básicas del método de elemento finito mostradas en la sección anterior.

Consideremos, entonces, una sección del material dividido en un cierto número de capas o elementos finitos con conductividad térmica constante. Las características de cada capa del material o elemento finito quedarán establecidas por la ley de Fourier. Llamaremos *odos*, para cada elemento finito, a los puntos de los planos que limitan a cada capa; claramente, la temperatura en cada nodo representará la temperatura uniforme en el plano correspondiente. Tomaremos, a modo de ejemplo, sólo tres elementos finitos (capas) y cuatro nodos (Fig. 5) que usualmente se denominan *odos globales*. Usaremos la notación (1), (2) y (3) para los elementos finitos y 1, 2, 3, y 4 para los nodos (lo correcto sería  $x_1, x_2, x_3$ , y  $x_4$ , pero por simplicidad se emplea generalmente esta notación). Como antes, consideremos un elemento típico. En este elemento (Fig. 6) los nodos 1 y 2 son *odos locales*,  $k^{(e)}$  es la conductividad térmica del elemento ( $e$ ) y  $q_1^{(e)}$  y  $q_2^{(e)}$  son las densidades de flujo de calor que entran y salen a los nodos 1 y 2. Por la ley de Fourier se sigue que

$$q_1^{(e)} = \frac{k^{(e)}}{L^{(e)}} (T_1^{(e)} - T_2^{(e)}),$$

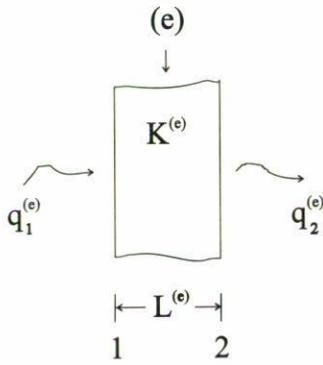


FIGURA 6. Elemento típico del conjunto con que se discretizó la pared.

con  $T_1^{(e)}$  y  $T_2^{(e)}$  las temperaturas en los nodos 1 y 2, respectivamente. Por conservación de energía

$$q_2^{(e)} = -q_1^{(e)},$$

por lo que

$$q_2^{(e)} = -\frac{k^{(e)}}{L^{(e)}} (T_1^{(e)} - T_2^{(e)}),$$

en donde hemos empleado la convención de que el calor añadido es positivo y el rechazado negativo ( $T_1^{(e)} > T_2^{(e)}$ ).

Si las dos últimas ecuaciones las escribimos en forma matricial, se tiene

$$\begin{bmatrix} \frac{k^{(e)}}{L^{(e)}} & -\frac{k^{(e)}}{L^{(e)}} \\ -\frac{k^{(e)}}{L^{(e)}} & \frac{k^{(e)}}{L^{(e)}} \end{bmatrix} \begin{bmatrix} T_1^{(e)} \\ T_2^{(e)} \end{bmatrix} = \begin{bmatrix} q_1^{(e)} \\ q_2^{(e)} \end{bmatrix}$$

para  $(e) = (1), (2)$  y  $(3)$ . O bien,

$$[k^{(e)}] [T^{(e)}] = [q^{(e)}]; \quad (k_{ij}^{(e)}) (T_i^{(e)}) = (q_i^{(e)}),$$

donde  $[k^{(e)}]$  es la matriz de coeficientes de conductividad térmica o *matriz de rigidez* (este nombre proviene de un problema de estructuras, que fue en donde se aplicó el método de elemento finito por vez primera),  $[T^{(e)}]$  es el vector columna de temperaturas nodales y  $[q^{(e)}]$  el vector columna de densidad de flujo de calor nodal. Notemos que la matriz  $[k^{(e)}]$  es simétrica.

Ahora procedamos a ensamblar los elementos finitos. Para llevar a cabo el ensamblaje proponemos construir la *topología del sistema*; es decir, realizar una identificación y registro de cuáles son los nodos globales que pertenecen a cada uno de los elementos y cómo se corresponden con los nodos locales. Como vemos, la topología del sistema es la que nos indicará la forma en que se van a unir los elementos manteniendo la continuidad. El ejemplo que hemos escogido hace que la topología sea muy sencilla, pues sólo basta establecer una correspondencia entre los nodos locales y globales, como

TABLA I. Tabla de correspondencia entre nodos locales y globales para el problema mostrado en la Fig. 5

Elemento	Nodo Local	Nodo Global
(1)	1	1
	2	2
(2)	1	2
	2	3
(3)	1	3
	2	4

se describe en la Tabla I. De esta tabla podemos establecer:

$$\begin{aligned} T_1^{(1)} &= T_1, & T_2^{(1)} &= T_2 = T_1^{(2)}, \\ T_2^{(2)} &= T_3 = T_1^{(3)}, & T_2^{(3)} &= T_4, \end{aligned}$$

y en cuanto a los demás términos:

$$k^{(1)} = k_1, \quad k^{(2)} = k_2, \quad k^{(3)} = k_3$$

y

$$L^{(1)} = L_1, \quad L^{(2)} = L_2, \quad L^{(3)} = L_3.$$

El ensamblaje se realizará con una aproximación directa o física. Calculemos, entonces, la densidad de flujo de calor neto que entra a cada uno de los nodos globales:

$$\text{Nodo1 : } q = \frac{k_1}{L_1} (T_1 - T_2)$$

$$\text{Nodo2 : } 0 = \frac{k_2}{L_2} (T_2 - T_3) - \frac{k_1}{L_1} (T_1 - T_2)$$

$$\text{Nodo3 : } 0 = \frac{k_3}{L_3} (T_3 - T_4) - \frac{k_2}{L_2} (T_2 - T_3)$$

$$\text{Nodo4 : } q = -\frac{k_3}{L_3} (T_3 - T_4).$$

Para los nodos 1 y 4 se adopta la convención usual de que calor añadido es positivo y rechazado negativo. En cuanto a los nodos 2 y 3, hay que tomar en cuenta que la densidad de flujo de calor neto es cero. Podemos escribir estas últimas ecuaciones en forma matricial:

$$\begin{bmatrix} \frac{k_1}{L_1} & -\frac{k_1}{L_1} & 0 & 0 \\ -\frac{k_1}{L_1} & \frac{k_1}{L_1} + \frac{k_2}{L_2} & -\frac{k_2}{L_2} & 0 \\ 0 & -\frac{k_2}{L_2} & \frac{k_2}{L_2} + \frac{k_3}{L_3} & -\frac{k_3}{L_3} \\ 0 & 0 & -\frac{k_3}{L_3} & \frac{k_3}{L_3} \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} q \\ 0 \\ 0 \\ -q \end{bmatrix}$$

o bien:

$$[k][T] = [q].$$

A la matriz  $[k]$  se le conoce como *matriz de rigidez ensamblada*.

Aun cuando las matrices de rigidez por elementos y la ensamblada sean singulares (no podemos encontrar su inversa, pues su determinante es cero), fácilmente podemos llegar a las siguientes ecuaciones:

$$T_2 = T_1 - q \frac{L_1}{k_1}; \quad T_3 = T_2 - q \frac{L_2}{k_2} = T_4 + q \frac{L_3}{k_3},$$

con

$$q = \frac{T_1 - T_4}{\frac{k_1}{L_1} + \frac{k_2}{L_2} + \frac{k_3}{L_3}},$$

que son las ecuaciones del flujo de calor para un muro multilaminar [8]. Para conocer las temperaturas intermedias ( $T_2$  y  $T_3$ ) basta conocer los valores:  $T_1, T_4, k_1, k_2, k_3, L_1, L_2$  y  $L_3$ ; después es posible determinar la densidad de flujo de calor.

El sistema de ecuaciones no podrá resolverse mientras la matriz  $[k]$  sea singular, pero esta matriz puede ser modificada si tomamos en cuenta las condiciones de frontera. Antes de realizar esto, es conveniente revisar retrospectivamente el proceso de ensamblaje.

La matriz de rigidez ensamblada o global contiene, como podemos percatarnos, entradas que pueden obtenerse sumando las entradas correspondientes a las matrices de rigidez de cada elemento, de acuerdo a una adecuada colocación en la matriz de rigidez global. Esta idea surge de una consideración física: el vector de carga resultante en un sistema de resortes lineales se puede obtener al sumar las cargas de los elementos individuales con una localización apropiada en la matriz columna.

Esto sugiere (recordemos cómo se obtuvo el área aproximada del círculo) que las matrices de rigidez de los elementos las podemos concebir como submatrices de la matriz de rigidez global, poniendo ceros en las entradas que no estén ocupadas por la submatriz. De manera que la matriz de rigidez global se puede obtener simplemente sumando las matrices así construidas.

Para nuestro ejemplo, las matrices de cada elemento serán submatrices de una matriz de  $4 \times 4$ , que es el orden de la matriz global. El mismo procedimiento se aplica al vector columna de densidad de flujo de calor, es decir, se considerará como submatriz de una matriz de  $4 \times 1$ , añadiendo ceros en las entradas extras. Tenemos que considerar, entonces, las siguientes matrices:

$$[K^{(1)}] = \begin{bmatrix} \frac{k_1}{L_1} & -\frac{k_1}{L_1} & 0 & 0 \\ -\frac{k_1}{L_1} & \frac{k_1}{L_1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad [q^{(1)}] = \begin{bmatrix} q \\ -q \\ 0 \\ 0 \end{bmatrix};$$

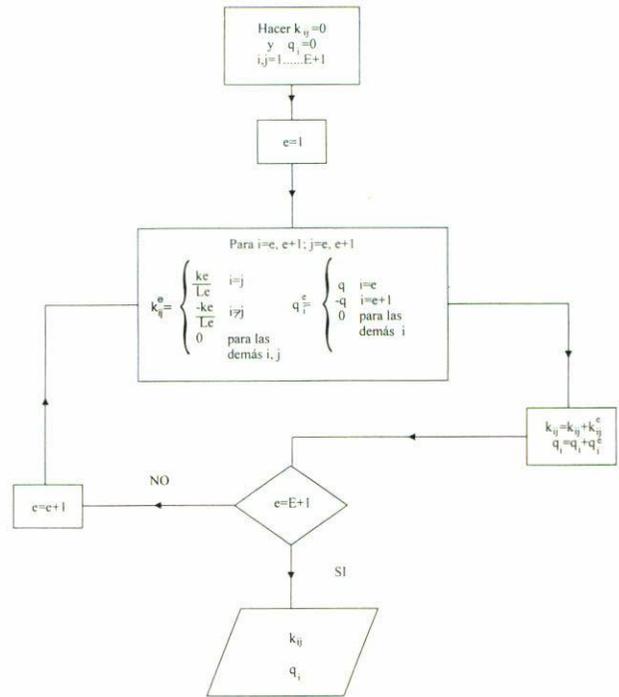


FIGURA 7. Diagrama de flujo para obtener las matrices globales ensambladas.

$$[k^{(2)}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{k_2}{L_2} & -\frac{k_2}{L_2} & 0 \\ 0 & -\frac{k_2}{L_2} & \frac{k_2}{L_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad [q^{(2)}] = \begin{bmatrix} 0 \\ q \\ -q \\ 0 \end{bmatrix};$$

$$[k^{(3)}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{k_3}{L_3} & -\frac{k_3}{L_3} \\ 0 & 0 & -\frac{k_3}{L_3} & \frac{k_3}{L_3} \end{bmatrix}; \quad [q^{(3)}] = \begin{bmatrix} 0 \\ 0 \\ q \\ -q \end{bmatrix}.$$

Sumando cada una de éstas, obtenemos

$$[k] = \sum_{e=1}^3 [k^{(e)}]; \quad [q] = \sum_{e=1}^3 [q^{(e)}].$$

Este proceso muestra la esencia del algoritmo que se emplea para realizar ensamblajes en el método de elemento finito (ver Fig. 7). Desde el punto de vista computacional, este algoritmo debe modificarse si se tienen limitaciones de memoria, pues el número de ceros que se manejan es alto. En general, si el número de capas en consideración fuera  $E$ , tendríamos que elegir  $E + 1$  nodos y la matriz de rigidez global, junto con la matriz de densidad de flujo de calor, puede obtenerse con el algoritmo mostrado en la Fig. 7.

Ahora bien, la matriz  $[k]$  es singular y no puede invertirse. La inclusión de las condiciones de frontera resuelve este problema. Al incluir las condiciones de frontera es conveniente notar que las que se refieren al flujo de calor en cada uno de los nodos ( $-k dT/dx|_{x=x_0}$ ) ya han sido incluidas durante el ensamblaje (están explícitas en la matriz  $[q] : q = -k dT/dx|_{x=x_1} = k dT/dx|_{x=x_4}$ ). A este tipo de condiciones se les llama *condiciones de frontera naturales o de Newman*, y a las variables involucradas en este tipo de condiciones se les llama *variables secundarias*. Las condiciones de frontera que especifican la temperatura en los nodos no están incluidas en el proceso de ensamblaje y deben especificarse. A estas condiciones, por tanto, se les llama *condiciones de frontera esenciales o de Dirichlet*, y a las variables, *variables primarias*.

Supongamos que tenemos especificados los valores en la frontera de la temperatura sólo en los extremos:  $T(x_1) = T_1$  y  $T(x_4) = T_4$ . Para ejemplificar, consideremos el sistema

$$\begin{bmatrix} k_{11} & k_{12} & 0 & 0 \\ k_{21} & k_{22} & k_{23} & 0 \\ 0 & k_{32} & k_{33} & k_{34} \\ 0 & 0 & k_{43} & k_{44} \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} q \\ 0 \\ 0 \\ -q \end{bmatrix}$$

Desarrollando, obtenemos las siguientes ecuaciones:

$$\begin{aligned} \frac{k_1}{L_1} (T_1 - T_2) &= q, \\ k_{22}T_2 + k_{23}T_3 &= -k_{21}T_1, \\ k_{32}T_2 + k_{33}T_4 &= -k_{34}T_4, \\ \frac{k_3}{L_3} (T_3 - T_4) &= q. \end{aligned}$$

Las variables primarias ( $T_2$  y  $T_3$ ), que son incógnitas, las podemos obtener de la segunda y tercera ecuación del sistema anterior. Las variables secundarias ( $q$ ) de la primera y cuarta, después de haber determinado las variables primarias desconocidas.

Podemos reescribir el sistema de manera convenientemente en forma matricial:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & k_{22} & k_{23} & 0 \\ 0 & k_{32} & k_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} T_1 \\ -k_{21}T_1 \\ -k_{34}T_4 \\ T_4 \end{bmatrix}$$

La nueva matriz  $[k]$  de este sistema es ahora invertible, por lo que es posible resolver el sistema y determinar las variables primarias y, con ellas, las variables secundarias. La conveniencia de escribir el sistema de este modo es que nos ayuda a especificar las condiciones de frontera esenciales cuando se presenten sistemas de  $n$  ecuaciones con  $n$  incógnitas. En efecto, supongamos que tenemos las matrices  $[k_{ij}]$  y  $[q_i]$  de orden  $n \times n$  y  $n \times 1$ , respectivamente. Para la especificación de las condiciones de frontera, procedemos como sigue [9]: si  $i$  es el subíndice correspondiente de una variable nodal prescri-

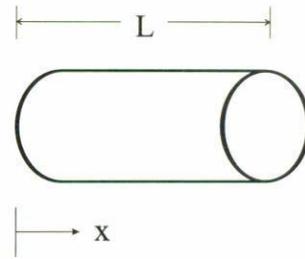


FIGURA 8. Barra aislada en donde se estudiará la transferencia de calor a lo largo de  $x$ .

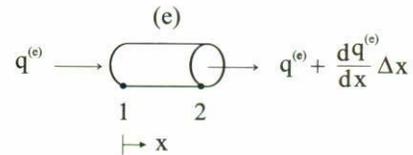


FIGURA 9. Balance energético en un elemento típico de la barra.

ta, el  $i$ -ésimo renglón y la  $i$ -ésima columna de  $[k]$  se igualan a cero, y  $k_{ii} = 1$ . El término  $q_i$  del vector columna  $[q]$  se reemplaza por el valor conocido  $T_i$ . Cada uno de los  $n - 1$  términos de  $[q]$  se modifican al restar de este término el valor de la variable prescrita, multiplicada por el término de la columna apropiada de la matriz original  $[k]$ . Este procedimiento se repite para cada uno de los valores prescritos hasta que todos ellos sean incluidos.

Una parte medular del método del elemento finito es obtener un sistema de ecuaciones como el anterior. Pero, como podemos percatarnos, hasta el momento sólo hemos discretizado el dominio en elementos finitos, falta todavía obtener la aproximación de la función que es la solución al problema que nos interesa. Esta aproximación a la solución se presenta mediante funciones, que usualmente son polinomios, que la interpolan localmente en los elementos finitos y, después de ensamblar, lo hacen globalmente. Por lo que, además de discretizar el dominio, hay que aproximar a la función solución mediante polinomios llamados, por esta razón, *polinomios de interpolación*.

### 3.1. Balance energético para un elemento finito

Consideremos ahora el flujo unidimensional de calor a través de una barra aislada que posee una fuente interna de calor (como una resistencia eléctrica). Supondremos que la barra, de longitud  $L$ , tiene sección transversal unitaria y que la temperatura es uniforme sobre cada sección transversal (Fig. 8).

Dividamos la barra en elementos finitos (por la uniformidad de la barra, basta conocer la temperatura en un punto, a saber, el nodo) y consideremos sólo uno de ellos (Fig. 9). La longitud del elemento ( $e$ ) es  $\Delta x$ . Representaremos por  $T^{(e)}$  y  $U^{(e)}$  a la distribución de temperatura y a la energía interna,

respectivamente. La razón a la que se genera calor, por unidad de longitud y unidad de tiempo, se representa por  $f(x)$  ó  $f^{(e)}$  para el elemento  $(e)$ , y  $q^{(e)}$  representa el flujo de calor en la dirección  $x$ , por unidad de área y unidad de tiempo, a través del elemento. Los cambios que ocurren en el elemento  $(e)$  durante un intervalo de tiempo  $\Delta t$ , es el incremento neto de calor ( $\Delta Q^{(e)}$ ):

$$\Delta Q^{(e)} = \text{flujo de calor que entra} - \text{flujo de calor que sale} \\ + \text{calor generado,}$$

es decir,

$$\Delta Q^{(e)} = q^{(e)} \Delta t - \left( q^{(e)} + \frac{dq^{(e)}}{dt} \Delta x \right) \Delta t + f^{(e)} \Delta x \Delta t,$$

$$\Delta Q^{(e)} = - \frac{dq^{(e)}}{dt} \Delta x \Delta t + f^{(e)} \Delta x \Delta t.$$

Para un sólido, la energía interna es una función continua de la masa. De ahí que exista una función  $\epsilon$ , llamada energía interna específica, tal que

$$U = \int_V \rho \epsilon dV,$$

donde  $\rho$  es la densidad del material y  $V$  el volumen. De esta ecuación obtenemos que el incremento de energía interna, por intervalo de tiempo, en el elemento  $(e)$  es

$$\dot{U} = \rho^{(e)} \dot{\epsilon}^{(e)} \Delta x.$$

Dado que no se realiza trabajo, la primera ley de la termodinámica establece que

$$\Delta \dot{Q} = \dot{U},$$

así

$$\rho^{(e)} \dot{\epsilon}^{(e)} = - \frac{dq^{(e)}}{dt} + f^{(e)}.$$

Como en los sólidos se cumple que  $\dot{\epsilon}^{(e)} = c_V^{(e)} dT^{(e)} / dt$ , siendo  $c_V^{(e)}$  el calor específico a volumen constante, entonces

$$- \frac{dq^{(e)}}{dt} + f^{(e)} = \rho^{(e)} c_V^{(e)} \frac{dT^{(e)}}{dt},$$

para  $x_1 \leq x \leq x_2$ . Por simplicidad en la exposición, hagamos una suposición más: la distribución de temperatura es estacionaria, esto es,  $dT/dt = 0$ . Con esto

$$- \frac{dq^{(e)}}{dt} + f^{(e)} = 0. \quad (2)$$

Esta relación que hemos encontrado se cumple sólo en puntos del elemento  $(e)$ ; para obtener una relación que se cumpla en la región que comprende el elemento tendríamos que integrar sobre todo el elemento. Antes de esto, y con el fin de

incluir  $T^{(e)}$  explícitamente en la ecuación, multipliquemos (2) por  $T^{(e)}$ :

$$- T^{(e)} \frac{dq^{(e)}}{dt} + T^{(e)} f^{(e)} = 0. \quad (3)$$

Como

$$- \frac{d}{dx} (T^{(e)} q^{(e)}) = T^{(e)} \frac{dq^{(e)}}{dx} + q^{(e)} \frac{dT^{(e)}}{dx},$$

entonces

$$- \frac{d}{dx} (T^{(e)} q^{(e)}) + q^{(e)} \frac{dT^{(e)}}{dx} + T^{(e)} f^{(e)} = 0.$$

Integrando sobre todo el elemento:

$$\int_{(e)} \left[ - \frac{d}{dx} (T^{(e)} q^{(e)}) + q^{(e)} \frac{dT^{(e)}}{dx} + T^{(e)} f^{(e)} \right] dx = 0, \quad (4)$$

que representa una relación que se cumple sobre una región finita, a saber, el elemento  $(e)$ .

Otro aspecto fundamental del método del elemento finito es que permite aproximar localmente, sobre cada elemento  $(e)$ , la temperatura, por medio de funciones sencillas que se derivan generalmente de la teoría de interpolación. Denotemos a los valores de la temperatura en los nodos locales por  $T^{(e)}(x_1) = T_1^{(e)}$  y  $T^{(e)}(x_2) = T_2^{(e)}$  que, aunque desconocidos, se interpolarán. Dado que se cuenta con dos puntos por donde la función debe pasar, la teoría de interpolación asegura que existe un único polinomio de primer grado que pasa por estos dos puntos. En efecto, elijamos una aproximación lineal de la forma

$$T^{(e)}(x) = ax + b.$$

Como hemos impuesto las condiciones esenciales de frontera para el elemento  $(e)$ :

$$T^{(e)}(x_1) = ax_1 + b = T_1^{(e)}, \\ T^{(e)}(x_2) = ax_2 + b = T_2^{(e)}, \quad (5)$$

que en forma matricial queda

$$\begin{bmatrix} T_1^{(e)} \\ T_2^{(e)} \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix},$$

de donde se sigue que

$$a = \frac{T_1^{(e)} x_2 - T_2^{(e)} x_1}{x_2 - x_1}; \quad b = \frac{T_2^{(e)} - T_1^{(e)}}{x_2 - x_1}.$$

Sustituyendo en (5) resulta

$$T^{(e)}(x) = \frac{T_1^{(e)} x_2 - T_2^{(e)} x_1}{x_2 - x_1} x + \frac{T_2^{(e)} - T_1^{(e)}}{x_2 - x_1} \\ = \frac{x_2 - x}{x_2 - x_1} T_1^{(e)} + \frac{x - x_1}{x_2 - x_1} T_2^{(e)},$$

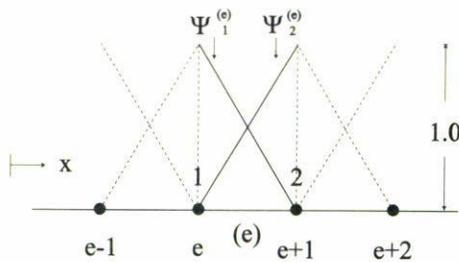


FIGURA 10. Funciones de aproximación locales.

o bien

$$T^{(e)}(x) = \sum_{i=1}^2 T_i^{(e)} \Psi_i^{(e)}. \tag{6}$$

A las funciones

$$\Psi_1^{(e)} = \frac{x_2 - x}{x_2 - x_1} \quad \text{y} \quad \Psi_2^{(e)} = \frac{x - x_1}{x_2 - x_1}, \tag{7}$$

se les denomina *funciones de interpolación locales* o *funciones de forma* (Fig. 10).

Además de que  $\Psi_i^{(e)} = 0$  fuera del elemento  $(e)$ , estas funciones de interpolación tienen las siguientes propiedades (Fig. 10):

- I.  $\Psi_i^{(e)}(x_j) = \begin{cases} 0, & \text{si } i \neq j; \\ 1, & \text{si } i = j. \end{cases}$
- II.  $\sum_{i=1}^2 \Psi_i^{(e)}(x) = 1.$

La propiedad I nos dice que (3) satisface las condiciones de frontera esenciales del elemento y II considera la posibilidad de que la temperatura sea constante en un intervalo (elemento). Es conveniente señalar que las *funciones de interpolación sólo dependen del tipo de elemento en consideración y no del problema que se está tratando.*

Sustituyendo (6) en la fórmula integral (4) se obtiene

$$\sum_{i=1}^2 T_i^{(e)} \int_{(e)} \left( -\frac{d}{dx} \Psi_i^{(e)} q^{(e)} + q^{(e)} \frac{d\Psi_i^{(e)}}{dx} + \Psi_i^{(e)} f^{(e)} \right) dx = 0.$$

Dado que esta expresión debe cumplirse para  $T_i^{(e)}$  arbitraria, se llega a que

$$\int_{(e)} q^{(e)} \frac{d\Psi_i^{(e)}}{dx} dx = - \int_{(e)} \Psi_i^{(e)} f^{(e)} dx + \left[ \Psi_i^{(e)} q^{(e)} \right]_{x_1}^{x_2}, \quad (i = 1, 2),$$

y haciendo la evaluación:

$$\int_{(e)} q^{(e)} \frac{d\Psi_i^{(e)}}{dx} dx = - \int_{(e)} \Psi_i^{(e)} f^{(e)} dx + \Psi_i^{(e)}(x_1) q_1^{(e)} - \Psi_i^{(e)}(x_2) q_2^{(e)} + q_i^{(e)},$$

para  $i = 1, 2$ . Aquí  $q_1^{(e)} = q^{(e)}(x_1)$  y  $q_2^{(e)} = q^{(e)}(x_2)$ , resultan ser las condiciones naturales de frontera del elemento.

Es importante notar que en todo este análisis no hemos tomado en cuenta el material con que está hecha la barra. Se puede caracterizar utilizando una relación constitutiva para la densidad de flujo de calor  $q^{(e)}$  dentro del elemento. La más común es la ley de Fourier:

$$q^{(e)} = -k^{(e)} \frac{dT^{(e)}}{dx}.$$

La conductividad térmica  $k^{(e)}$  la supondremos constante en todo el elemento  $(e)$ .

Usando nuevamente la aproximación local para  $T^{(e)}$  (Ec. 6):

$$q^{(e)} = -k^{(e)} \sum_{j=1}^2 T_j^{(e)} \frac{d\Psi_j^{(e)}}{dx},$$

y sustituyendo en (4):

$$\sum_{i=1}^2 \int_{(e)} \left( k^{(e)} \frac{d\Psi_i^{(e)}}{dx} \frac{d\Psi_j^{(e)}}{dx} \right) dx = \int_{(e)} \Psi_i^{(e)} f^{(e)} dx + q_i^{(e)}, \quad (i = 1, 2).$$

Que se puede escribir como:

$$\sum_{i=1}^2 K_{ij}^{(e)} = F_i^{(e)} + q_i^{(e)} = f_i^{(e)},$$

donde

$$K_{ij}^{(e)} = \int_{(e)} \left( k^{(e)} \frac{d\Psi_i^{(e)}}{dx} \frac{d\Psi_j^{(e)}}{dx} \right) dx, \\ F_i^{(e)} = \int_{(e)} \Psi_i^{(e)} f^{(e)} dx,$$

para  $i, j = 1, 2$ . Ahora, si integramos la primera de estas expresiones ( $\Psi_1^{(e)} = 1 - \bar{x}/h_e$  y  $\Psi_2^{(e)} = \bar{x}/h_e$ , con  $0 \leq \bar{x} \leq h_e$ ):

$$K_{ij}^{(e)} = \begin{cases} \frac{k^{(e)}}{h_e} & i = j \\ -\frac{k^{(e)}}{h_e} & i \neq j \end{cases}$$

con  $h_e = x_2 - x_1$ . El sistema resultante puede escribirse en forma matricial como

$$\begin{bmatrix} \frac{k^{(e)}}{h_e} & -\frac{k^{(e)}}{h_e} \\ \frac{k^{(e)}}{h_e} & \frac{k^{(e)}}{h_e} \end{bmatrix} \begin{bmatrix} T_1^{(e)} \\ T_2^{(e)} \end{bmatrix} = \begin{bmatrix} f_1^{(e)} \\ f_2^{(e)} \end{bmatrix} \begin{bmatrix} F_1^{(e)} \\ F_2^{(e)} \end{bmatrix} + \begin{bmatrix} q_1^{(e)} \\ q_2^{(e)} \end{bmatrix}$$

El último paso es el ensamblaje, el cual se realiza como se describió en la primera parte de este trabajo. Si  $E$  es el número de elementos en consideración, para cada elemento se suman submatrices de orden  $(E + 1) \times (E + 1)$  y vectores columna de orden  $(E + 1) \times 1$ . La imposición de condiciones de frontera se realiza también en forma similar a la ya expuesta anteriormente.

Después del proceso de ensamblaje obtenemos los valores  $T_i = T(x_i)$ ,  $i = 1, \dots, E + 1$ , que son los valores de la temperatura en cada uno de los nodos globales. Aún más, por las propiedades de las funciones de forma, podemos obtener la solución aproximada  $\bar{T}$  para todo el dominio:

$$T \simeq \bar{T}_h(x) = \sum_{e=1}^E \left( \sum_{i=1}^2 T_i^{(e)} \Psi_i^{(e)} \right) = \sum_{j=1}^N T_j \Phi_j(x), \quad (8)$$

donde  $\Phi_J(x)$ ,  $J = 1, \dots, N$  ( $N = E + 1$ ), son las llamadas *funciones de interpolación global*:

$$\Phi_J(x) = \begin{cases} \Psi_2^{(J-1)}(x) & x_{J-1} \leq x \leq x_J \\ \Psi_1^{(J)}(x) & x_J \leq x \leq x_{J+1} \end{cases}$$

## 4. Algunos aspectos matemáticos del método del elemento finito

### 4.1. Planteamiento del problema

Regresemos nuevamente al problema de hallar la distribución de temperatura  $T(x)$  en una barra de longitud  $L$ . Consideremos por simplicidad  $L = 1$  y con la intención de simplificar la notación, denotaremos a la temperatura por  $u(x)$  en lugar de  $T(x)$ , de tal suerte que nuestro problema se reduce a hallar  $u(x)$  tal que

$$\begin{aligned} -k \frac{d^2 u}{dx^2} &= f, & 0 \leq x \leq 1, \\ u(0) &= 0; & u(1) = 0. \end{aligned} \quad (9)$$

La solución  $u(x)$  a este problema debe tener, al menos, segunda derivada continua, propiedad que también debe cumplir cualquier solución aproximada  $\hat{u}(x)$  a este problema. En lo que se refiere al método del elemento finito, las restricciones de continuidad de la solución se reflejan en limitaciones de, por ejemplo, la clase de elementos que se pueden utilizar para discretizar el dominio y en el tamaño de las matrices involucradas. Por otro lado, en el caso del problema (9), podemos ver que buscar una solución con primera derivada continua es una restricción muy severa.

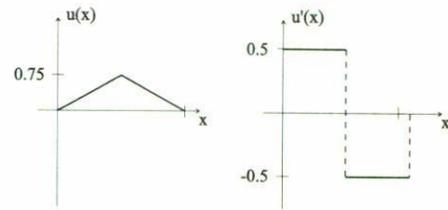


FIGURA 11. Solución al problema (10) y su primera derivada.

La dificultad se encuentra en la “suavidad” del término no-homogéneo de (9), que constituye la información o los “datos” del problema. Como ejemplo, consideremos el siguiente problema:

$$\begin{aligned} -k \frac{d^2 u}{dx^2} &= \delta \left( x - \frac{1}{2} \right), & 0 \leq x \leq 1, \\ u(0) &= 0; & u(1) = 0. \end{aligned} \quad (10)$$

La cantidad  $\delta \left( x - \frac{1}{2} \right)$ , la delta de Dirac, puede interpretarse como un pulso unitario de corriente en la fuente interna de calor. La solución de (10) es

$$u(x) = \begin{cases} \frac{x}{2}, & 0 \leq x \leq \frac{1}{2} \\ \frac{1}{2}(1-x) & \frac{1}{2} \leq x \leq 1. \end{cases}$$

La gráfica de esta expresión, junto con la de su primera derivada se muestran en la Fig. 11.

En la figura observamos que la solución  $u(x)$  de (10) presenta en su primera derivada un salto unitario en  $x = 1/2$ , por lo que  $u''(x)$  no existe. ¿Cómo puede entonces una función  $u(x)$  ser solución al problema (10), en  $0 \leq x \leq 1$ , si su segunda derivada no existe en todo el intervalo, debido a lo irregular de los datos del problema?

Esta clase de dificultades pueden evitarse en algunos casos formulando el problema (9) de una manera alternativa llamada *formulación débil o variacional*. Es importante señalar que esta formulación no siempre es posible de obtener. En nuestro ejemplo es posible, como veremos enseguida.

### 4.2. Formulación variacional del problema

La idea principal consiste en llevar el problema de un planteamiento diferencial (9) a uno integral, como se hizo en la sección anterior. Esto se consigue multiplicando (9) por una función  $v(x)$ , llamada *función de prueba*, e integrando sobre todo el dominio, de manera que el problema se transforma en hallar  $u(x)$  tal que

$$\begin{aligned} \int_0^1 \left( -k \frac{d^2 u}{dx^2} \right) v \, dx &= \int_0^1 f v \, dx, & 0 \leq x \leq 1 \\ u(0) &= 0; & u(1) = 0, \end{aligned}$$

que debe cumplirse para toda  $v(x)$ , tal que  $v(0) = v(1) = 0$ . Integrando por partes (fórmula de Green) y tomando en cuenta que  $v = 0$  en los extremos, obtenemos un planteamiento más simétrico del problema: hallar  $u(x)$  tal que

$$\int_0^1 (k u'v' - fv) dx = 0, \quad 0 \leq x \leq 1$$

$$u(0) = u(1) = 0, \quad v(0) = v(1) = 0.$$

Podemos notar que ahora  $u(x)$  necesita tener sólo primera derivada continua [lo mismo ocurre con  $v(x)$ ]. A cambio, ahora debemos asegurarnos de que el primer integrando de (4.2) pueda evaluarse. Si tomamos  $v = u$ , esto se traduce en que las integrales del tipo  $\int_0^1 (u')^2 dx$  sean finitas, esto es,  $\int_0^1 (u')^2 dx < \infty$ . La clase de funciones que cumplen con esta restricción se dice que son funciones con primera derivada *cuadrado integrable*.

Esta clase de restricciones y, en general, el planteamiento del problema, se pueden escribir en una forma más compacta y útil si agrupamos las funciones en *clases*. Todas las funciones  $w$  con primera derivada cuadrado integrable y nulas en los extremos, se agruparán en una clase que llamaremos  $H$ , de tal suerte que  $u \in H$  si  $\int_0^1 (w')^2 dx < \infty$  y  $w(0) = w(1) = 0$ . De otra manera, la clase  $H$  se define como

$$H = \left\{ w \mid \int_0^1 (w')^2 dx < \infty, \quad w(0) = w(1) = 0 \right\}.$$

Con esto, la formulación variacional del problema (9) se presenta de la siguiente manera

Hallar  $u \in H$  tal que

$$\int_0^1 (ku'v' - fv) dx = 0, \quad \forall v \in H. \quad (11)$$

La clase  $H$  se denomina *clase de funciones admisibles* de (11).

Es interesante notar que pasar de (9) a (11) ha implicado debilitar progresivamente los requerimientos de suavidad de la solución y, por tanto, aumentar la clase de datos para los cuales nuestro problema tiene sentido.

Es posible conocer la naturaleza de la clase de funciones  $H$  en un contexto más general, en el sentido de que sea igualmente válido para problemas diferentes a (9), como pueden ser problemas de transporte o de mecánica de sólidos. Generalmente  $H$  se identifica con una clase bien conocida de funciones llamada *clase de Sobolev*, y denotada por  $H^m(\Omega)$ , que se define como sigue.

- Supongamos que  $\Omega$  es una región acotada en  $R^3$  (en el caso del problema (9),  $\Omega$  es una región en  $R^1$  dada por  $0 \leq x \leq 1$ ) y que  $u = u(x, y, z)$  es una función de posición en  $\Omega$ . Entonces  $u \in H^m(\Omega)$  si  $u$  y todas sus derivadas parciales de orden menor o igual a  $m$  son cuadrado integrables. En notación compacta, para  $\Omega \subset R^2$ , escribimos

$$H^m(\Omega) = \left\{ u \mid u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \dots, \frac{\partial^m u}{\partial z^m} \in L^2(\Omega) \right\},$$

donde  $L^2(\Omega)$  se usó para denotar a la clase de funciones que son cuadrado integrables en  $\Omega$ .

Como ejemplo podemos considerar la función  $u(x)$  de la Fig. 11. Claramente,  $u'$  es discontinua en  $x = 1/2$ , sin embargo es cuadrado integrable, pues  $\int_0^1 (u')^2 dx = 1/4$ , por lo que  $u(x) \in H^1(0, 1)$ .

Es usual tomar también en cuenta las condiciones de frontera al definir las clase de Sobolev. Así, denotamos por  $H_0^m(\Omega)$  al espacio de funciones que están en  $H^m(\Omega)$  y que además

$$u = 0, \quad \frac{\partial u}{\partial n} = 0, \quad \frac{\partial^2 u}{\partial n^2} = 0, \quad \dots, \quad \frac{\partial^{m-1} u}{\partial n^{m-1}} = 0,$$

en la frontera de  $\Omega$ , donde  $n$  es el vector unitario normal que apunta hacia afuera de la frontera.

Y con esto completamos el formalismo básico para poder plantear de manera más concreta cualquier problema variacional con valores en la frontera. En particular, regresando al problema (9) y a su forma débil (11), podemos ver que la clase  $H$  de funciones admisibles en  $\Omega = (0, 1)$  es justamente la clase de Sobolev  $H_0^1(0, 1)$ , de manera que el planteamiento variacional de (9) de hallar  $u \in H_0^1(0, 1)$  tal que

$$\int_0^1 (ku'v' - fv) dx = 0, \quad \forall v \in H_0^1(0, 1). \quad (12)$$

El método del elemento finito que hemos aplicado, parte de este último planteamiento para hallar una solución aproximada a (9) siguiendo el procedimiento mostrado en la Sec. 3.

Aunque no se hizo explícito en ese momento, ahora podemos apreciar que al aplicar la fórmula de la derivada de un producto de funciones en (3) lo que conseguimos fue formular el problema, originalmente diferencial (2), en forma variacional (4).

Antes de considerar aspectos del método del elemento finito en  $H^m(\Omega)$  es ilustrativo revisar un método para obtener una solución aproximada de (12) que involucra un procedimiento que es justamente el punto de partida del método del elemento finito.

### 4.3. Aproximación de Galerkin

El método de Galerkin consiste en buscar una solución aproximada de (12), pero no en todo el espacio  $H_0^1$ , sino en un subespacio de dimensión finita  $N$ ,  $H^{(N)}$ , de  $H_0^1$  ( $H^{(N)} \subset H_0^1$ ). De manera que ahora el problema es hallar  $u_N \in H^{(N)}$ , tal que

$$\int_0^1 (ku'_N v'_N) dx = \int_0^1 (fv_N) dx, \quad \forall v_N \in H^{(N)}. \quad (13)$$

Ahora, si  $\{\phi_1, \phi_2, \dots, \phi_N\}$  es una base de  $H^{(N)}$  (esto es, son linealmente independientes y generan a  $H^{(N)}$ ), cualquier  $u_N \in H^{(N)}$  se puede escribir como

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

donde  $x \in [0, 1]$ . De manera que dada la base  $\{\phi_i\}$ , los coeficientes  $\alpha_i$  pueden determinarse sustituyendo la expresión anterior en (13).

Entonces, dados  $u_N$  y  $v_N$  de  $H_0^{(N)}$ , podemos escribir

$$u_N = \sum_{i=1}^N \alpha_i \phi_i(x), \quad y \quad v_N = \sum_{i=1}^N \beta_i \phi_i(x). \quad (14)$$

Y puesto que las  $\{\phi_i\}$  son conocidas,  $u_N$  estará completamente determinada una vez que los coeficientes  $\alpha_i$  sean calculados\*

Sustituyendo (14) en (13) se obtiene un sistema de  $N$  ecuaciones lineales para los  $N$  coeficientes  $\alpha_j$ :

$$\sum_{j=1}^N K_{ij} \alpha_j = F_i,$$

para  $i = 1, 2, \dots, N$ . O bien  $[K][\alpha] = [F]$ , donde  $[K]$  es la matriz de rigidez y al vector  $[F]$  se le denomina *vector de carga*.

Es importante notar que la calidad de la aproximación depende de la elección de las funciones  $\phi_i$ , ya que, una vez definidas, la determinación de los coeficientes  $\alpha_j$  es un problema meramente computacional.

#### 4.4. El método del elemento finito en $H^m(\Omega)$ .

El método de Galerkin provee una estrategia para construir soluciones a problemas con valores en la frontera pero, como puede apreciarse, tiene el inconveniente de que no incluye una forma sistemática para escoger las funciones de aproximación  $\phi_i(x)$ ; la elección es diferente para cada problema y, puesto que estas funciones deben satisfacer las condiciones de frontera, su elección resulta tanto más difícil cuanto más irregular sea la geometría. El método del elemento finito ofrece una forma general y sistemática para generar las funciones base  $\phi_i(x)$ .

Antes de mostrar cómo este método procede para generar estas funciones, es necesario introducir un nuevo cambio en la notación. En la formulación variacional (12) denotamos a las funciones de prueba por  $v_N$  y a la solución del problema por  $u_N$ , donde  $N$  es un parámetro que indica el número de funciones base que generan  $H_0^{(N)}$ . En el planteamiento matemático del método del elemento finito es costumbre usar la longitud  $h$  del elemento utilizado para discretizar el dominio como el parámetro que sustituye a  $N$ . La idea es que mientras más pequeño sea el valor de  $h$ , se necesitarán más funciones base para  $H_0^{(N)}$ . Así, en lo que sigue, denotaremos a  $v_N, u_N$  y  $H_0^{(N)}$  por  $v_h, u_h$  y  $H_0^{(h)}$ .

Como se ha mostrado a lo largo de este trabajo, el primer paso del método del elemento finito consiste en reemplazar el dominio  $\Omega$  de la solución al problema, por una colección  $\Omega_h$  de dominios simples  $\Omega_e$ : los elementos finitos. Esto es (Fig. 12),

$$\Omega \approx \Omega_h, \quad \bar{\Omega}_h = \sum_{e=1}^E \bar{\Omega}_e.$$

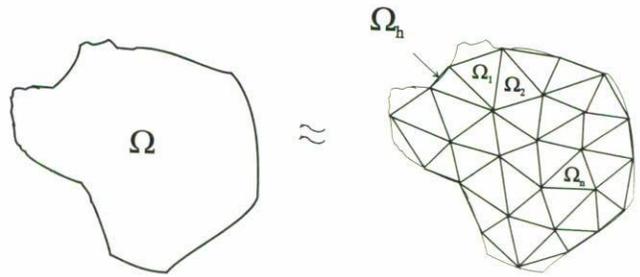


FIGURA 12. Discretización de un dominio arbitrario en elementos finitos.

Para las funciones base  $\phi_j$ , usualmente se eligen polinomios (principalmente porque son estables desde el punto de vista computacional) que se construyen de la siguiente manera:

- Sobre cada elemento finito  $\Omega_e$  se construye una *función de aproximación local*  $\Psi_i^e, i = 1, \dots, N_e$  (Ec. 7), de manera que la solución  $u_h$  en el elemento  $\Omega_e$  es (véase la Ec. 6):

$$u_h^{(e)} = \sum_{j=1}^{N_e} \alpha_j^e \Psi_j^{(e)}(x).$$

- Las *funciones globales de interpolación*  $\Phi_j$ , se obtienen uniendo topologicamente las funciones de aproximación locales como se mostró en la sección anterior (Ec. 8). Dado que las funciones locales tienen la propiedad de asumir valores 1 o 0 en los nodos (Fig. 10), es directo generalizar esta condición a cualquier derivada de la función de aproximación:

$$\frac{d^n \Psi_i^{(e)}(x_j^{(e)})}{dx_k^n} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases} \quad (15)$$

- Entonces, en un nodo global  $x_j \in \Omega_h$ , las funciones de interpolación globales  $\Phi_j(x)$  deben tener también la propiedad de que su valor, o el de sus derivadas, sea 1 o 0 en ese nodo. Como ejemplo, supongamos que cierto nodo  $i$  es compartido por  $E$  elementos finitos. Las funciones locales correspondientes a cada uno de estos elementos deben combinarse en tal forma que resulte una sola función global  $\Phi_j$  para ese nodo que satisfaga los requerimientos de continuidad entre elementos [ver Fig. 10 y Ec. (15)].

En analogía con el método de Galerkin, al generar las funciones de interpolación globales  $\Phi_j$ , obtenemos una base del espacio  $H^h$ , donde  $H^h \subset H^m(\Omega)$ . A los elementos finitos en donde se definen funciones  $\Phi_j$  con esta propiedad, se les denomina *elementos conformes*. Para determinar condiciones suficientes para que un elemento sea conforme, debemos observar que puesto que las funciones  $v_h$  son, por lo general, infinitamente regulares (polinomios) en el interior de todos los

elementos, la regularidad de las funciones en todo  $\Omega$  depende de hasta que orden de derivadas son continuas en la frontera entre elementos. Denotemos por  $C^n(\Omega)$  la clase de funciones con derivada continua de orden  $n$  en  $\Omega$ . Si  $v_h \in C^0(\Omega)$ , pero hay, por ejemplo, un escalón en la derivada de  $v_h$ , normal a la frontera entre elementos, entonces  $\partial^2 v_h / \partial n^2$  (donde  $n$  es el vector normal a la frontera del elemento) estará indefinida. En consecuencia,  $v_h \in H^1(\Omega)$  pero  $v_h \notin H^2(\Omega)$ . En general, si  $v_h$  es tal que sus derivadas de orden  $m-1$  son continuas en  $\Omega$ , esto es  $v_h \in H^{m-1}(\Omega)$ , entonces  $v_h \in H^m(\Omega)$  y tales elementos conformes en  $H^m(\Omega)$  se denominan *elementos*  $C^{m-1}$ .

#### 4.5. Errores de aproximación

El error de aproximación se define como

$$e(x) = u(x) - u_h(x),$$

donde  $u(x)$  es la solución exacta al problema y  $u_h(x)$  es la solución obtenida por el método del elemento finito. Es claro que con esta ecuación el error no puede calcularse a menos que se conozca la solución exacta. No obstante, si  $u(x)$  es desconocida (que es, generalmente, el caso), es posible hacer una estimación del error y determinar si este es menor cuando  $h$  (el tamaño de los elementos) es menor y, por tanto, el número de elementos es mayor.

Puesto que los errores son funciones y debemos medir su "tamaño", es necesario saber cómo se define la norma de una función en  $H^m(\Omega)$ . Ésta es

$$\|e\|_m = \left\{ \int_0^1 \left[ \left( \frac{d^m u}{dx^m} \right)^2 + \left( \frac{d^{m-1} u}{dx^{m-1}} \right)^2 + \dots + \left( \frac{du}{dx} \right)^2 + u^2 \right] dx \right\}^{1/2}.$$

Ahora bien, se espera que la solución se aproxime cada vez más a la exacta cuando se tomen elementos cada vez más pequeños (proceso que se conoce con el nombre de *refinamiento* de la malla), es natural esperar que una expresión para el error en  $H^m(\Omega)$  sea función del tamaño del elemento:

$$\|e\|_m = f(h).$$

Dado un problema con valores en la frontera y su solución aproximada  $u_h$  obtenida con el método de elemento finito, obtener  $\|e\|_m$  es un problema matemáticamente complicado que ha sido resuelto para casos muy particulares y que se complica aún más cuando se trabaja en 2 y 3 dimensiones. Con fines ilustrativos, sin embargo, podemos mostrar la forma típica del comportamiento del error en un problema unidimensional bajo ciertos requerimientos.

Supongamos que prevalecen las siguientes condiciones:

1. Se usan elementos conformes  $C^{m-1}$ , esto es,  $H^h \subset H^m(\Omega)$ .
2. En cada elemento, las funciones de aproximación local  $\Psi_i^e$  son polinomios completos de grado  $k$ .
3. La solución exacta  $u$  es tal que  $u(x) \in H^r(\Omega)$  para  $r > m$ .

En tal caso, se puede mostrar [1] que

$$\|e\|_m \leq Kh^\nu,$$

donde  $\nu$  es mínimo entre  $(k+1-m)$  y  $(r-m)$  y  $K$  es una constante que depende del tipo de problema y condiciones iniciales. De manera que, bajo las condiciones supuestas, el error está acotado y el método converge con una cierta velocidad que está dada por el exponente  $\nu$ .

\*. Notese la importancia de que ambas,  $u_N$  y  $v_N$ , se encuentren en  $H_0^{(N)}$ , pues sólo en este caso podemos representar a  $u_N$  y  $v_N$  en la misma base, como en (14). A este tipo de formulación se le denomina *formulación variacional simétrica*.

1. J.N. Reddy, *Applied Functional Analysis and Variational Methods in Engineering*, (McGraw Hill, New York, 1986).
2. K.H. Huebner and E.A. Thornton, *The Finite Element Method for Engineers*, 2nd edition (Wiley, New York, 1982).
3. J.T. Oden, *Finite Elements of Nonlinear Continua*, (McGraw Hill, New York, 1972).
4. E.B. Becker, G.F. Carey, and J.T. Oden, *Finite Element, an Introduction*, Vol. 1 (Prentice-Hall Inc., Englewood Cliffs, N.J., 1983).

5. J.N. Reddy, *An Introduction to the Finite Element Method*, (McGraw Hill, New York, 1984).
6. J.T. Oden and G.F. Carey, *Finite Elements, Mathematical Aspects*, (Prentice-Hall Inc., Englewood Cliffs, N.J., 1983).
7. R. Courant and H. Robbins, *What is Mathematics?*, (Oxford University Press, New York, 1953).
8. M.A. Mijeev y I.M. Mijeeva, *Fundamentos de Termodinámica*, (MIR, Moscu, 1977).
9. C.A. Felippa and R.W. Clough, *The Finite Element Method in Solid Mechanics*, Numerical Solutions of Field Problems in Continuum Physics, SIAM-AMS Proceedings, Vol. 2, (1970).