



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE MATEMÁTICA



**“Existencia y Unicidad de Soluciones
para la Ecuación de Convección -
Difusión - Reacción del Tipo
Estacionario Usando los Métodos de
Galerkin y Esquema Monótono.”**

Tesis

Para optar el título profesional de

LICENCIADA EN MATEMÁTICAS

Autor:

BACH.MAT. JUANA SMITH BANCES ACOSTA

Asesor:

DR. CÁRPENA VELÁSQUEZ ENRIQUE WILFREDO



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE MATEMÁTICA



Los firmantes, por la presente certifican que han leído y recomiendan a la Facultad de Ciencias Físicas y Matemáticas la aceptación de la tesis titulada “Existencia y Unicidad de Soluciones para la Ecuación de Convección - Difusión - Reacción del Tipo Estacionario Usando los Métodos de Galerkin y Esquema Monótono”, presentado por la Bach.Mat. Juana Smith Bances Acosta, en el cumplimiento parcial de los requisitos necesarios para la obtención del título profesional de Licenciada en Matemáticas

Dra. Gloria María Ortiz Basauri
Presidente del Jurado

Msc. Juana Doris Blas Rebaza
Secretario del Jurado

Mg. Betty Rimarachín López
Vocal del Jurado

Fecha de defensa: 10 de Setiembre del 2020



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE MATEMÁTICA



TESIS:

**EXISTENCIA Y UNICIDAD DE SOLUCIONES PARA LA
ECUACIÓN DE CONVECCIÓN - DIFUSIÓN - REACCIÓN DEL
TIPO ESTACIONARIO USANDO LOS MÉTODOS DE
GALERKIN Y ESQUEMA MONÓTONO.**

Bach.Mat. Juana Smith Bances Acosta

Autora de la tesis

Dr. Cárpene Velásquez Enrique Wilfredo

Asesor

Fecha de defensa: 10 de Setiembre del 2020

AGRADECIMIENTOS

En primer lugar agradecer a mis padres Agustín Bances y Rosa Acosta por ser los principales promotores e impulsores de mis sueños y proyectos alcanzados. Gracias a ellos por todos los días confiar en mi capacidad y expectativas; a mi madre por estar siempre dispuesta en cada momento que lo necesité, por ser mi amiga y compañera en toda mi vida; a mi padre por desear y anhelar lo mejor para mi vida, y por sus consejos que me guían y me hacen mejor cada día.

A mi bebé Néstor Rodrigo por ser mi gran motivación e impulsor de mi superación para ofrecerle siempre lo mejor. No es fácil, lo sé, pero sino te tuviese a mi lado no habría logrado tantas cosas. A mi pareja y padre de mi bebé, Nestor Sánchez, por su amor, comprensión, por siempre creer en mi capacidad, por sus consejos, tanto en lo personal como en lo académico, gracias por tu cariño corazón. A mi sobrino y ahijado, Antonio Agustín, a quien quiero y adoro como un hijo.

A mis hermanos Dalila, María, Agustín, que con sus palabras de aliento no me dejaban caer para que siga adelante, mantenerme perseverante y pueda cumplir mis metas, gracias por todos los momentos que hemos vivido juntos.

A mi tutor de tesis, el profesor Enrique Carpena, por la paciencia y tolerancia que tuvo durante la etapa de mis estudios de pregrado y en la elaboración de mi tesis para la titulación, por brindarme la oportunidad de trabajar con él en mi tesis, por los consejos y sugerencias que me permitieron crecer en el campo del Análisis Numérico. A su vez, quiero expresarle mi respeto y gratitud por ser un excelente profesor y guía de tesis, expresar también mi admiración, no sólo en lo profesional sino también en lo personal.

Éste logro es gracias a todos ustedes.

RESUMEN

En éste trabajo utilizamos un método de análisis numérico que sirve para aproximar la solución de ecuaciones diferenciales parciales con valores iniciales, llamado Método de Elementos Finitos (MEF), que está pensado para ser usado en computadoras y que proporciona una técnica particular sistemática para construir funciones bases como funciones polinomiales por tramos sobre subregiones del dominio, estas funciones son llamadas elementos finitos. Para ello, se empieza derivando la ecuación diferencial parcial con condiciones iniciales a un problema denominado problema débil, que consta de una forma bilineal y un funcional lineal. Además, a dicha forma bilineal se le asocia una matriz llamada matriz de rigidez.

El objetivo de éste trabajo es el análisis de la existencia y unicidad de las soluciones, continua y discreta, de la ecuación de convección-difusión-reacción del tipo estacionario. Para el análisis de las soluciones discretas se utilizan dos métodos de elementos finitos; el método de Galerkin clásico y un método Monótono. Las ecuaciones diferenciales parciales de convección-difusión-reacción del tipo estacionario modelan, por ejemplo, la concentración de partículas en un campo de flujo de velocidad externa. En particular, nos centraremos especialmente en las ecuaciones de convección-difusión-reacción con convección dominada, es decir, cuando el término difusivo es varios ordenes más pequeño que el término convectivo.

Debido a que el problema es de convección dominada los esquemas clásicos de elementos finitos no son eficientes para resolver este tipo de problemas, ya que la solución numérica tiende a oscilar si la medida de la malla no es suficientemente pequeña. Teniendo en cuenta ese inconveniente es que surgen los esquemas monótonos de elementos finitos, éstos esquemas monótonos tienen la característica particular de satisfacer el Principio del Máximo Débil tanto a nivel continuo como a nivel discreto, denominado propiedad de monotonía.

El método monótono que desarrollaremos en este trabajo es el Esquema Monótono de Elementos Finitos del Borde Promediado denominado EAFEM, donde la idea principal es promediar adecuadamente los coeficientes de la ecuación diferencial en las aristas de la triangulación o malla. En éste método, la matriz de rigidez resultante es una matriz llamada M-matriz bajo ciertas suposiciones sobre la triangulación (generalmente no estructurada).

El método fue propuesto por Xu y Zikatanov en [5] y motivado por los trabajos de Markowich y Zlamal [7] y Brezzi, Marini y Pietra [8]. Además, Xu Zikatanov en [5], nos dice que éste esquema (EAFEM), es un esquema monótono de upwinding, es decir, la solución numérica tiende a ser suave.

En el Capítulo 1, se presentan definiciones y resultados sobre la teoría de elementos finitos, Identidades de Green, Principio el Máximo Débil, esquema de Galerkin, esquema monótono, los cuales se utilizan para la demostraciones de la existencia y unicidad de las soluciones de la ecuación de convección–difusión–reacción del tipo estacionario y el estudio de las estimaciones de error *a priori*.

En el segundo capítulo, se aplican resultados para garantizar existencia y unicidad de la solución continua de la ecuación de convección–difusión–reacción del tipo estacionario, entre ellos, el teorema de Lax Milgram, y se corrobora que la ecuación diferencial parcial en estudio satisface la propiedad de monotonía a nivel continuo.

En el capítulo 3, se presenta el esquema discreto de Galerkin para el cual se utilizan elementos finitos lineales, continuos y a trozos, y se estudia el esquema EAFEM propuesto por Xu Zikatanov en [5]. Además de estudiar la existencia y unicidad de la solución discreta para ambos métodos, se detalla y corrobora las cuentas de la demostración del teorema del error de convergencia para el esquema discreto EAFEM. Finalmente, se presenta un resultado numérico que muestra el buen comportamiento del método EAFEM y la ineficiencia del Método de Galerkin.

INTRODUCCIÓN

Las ecuaciones de convección–difusión–reacción son utilizadas en varias aplicaciones de ingeniería ambiental e industrial. Un ejemplo de simulación de estas ecuaciones es el transporte de sustancias en medios acuáticos, utilizada para medir el impacto de una fuente de contaminantes en un río. Ésta dinámica de concentración de partículas pequeñas es determinado por un campo de flujo de velocidad externa. Otro tipo de simulación típica de éstas ecuaciones se da cuando el tamaño difusivo es varios órdenes de magnitud más pequeño que el término convectivo.

El tipo de ecuaciones que son interesante estudiar son las ecuaciones de convección–difusión–reacción con convección dominada, es decir, cuando el cociente entre el término convectivo y el término difusivo es demasiado grande. El método de elementos finitos clásico y el método de diferencias finitas, son en general, esquemas inestables para la solución de éste tipo de ecuaciones, en el sentido que la solución numérica frecuentemente contiene oscilaciones si la medida de la malla no es suficientemente pequeña.

Una de las formas de eliminar éstas oscilaciones es considerar métodos o esquemas monótonos de elementos finitos. Una característica para que un esquema sea monótono, es que el esquema satisfaga el Principio del Máximo Débil a nivel continuo, y que la ecuación discreta resultante también cumpla el Principio del Máximo Débil Discreto (que es similar al del nivel continuo).

El Principio del Máximo Débil es una propiedad típica cualitativa de problemas lineales de valores de frontera elíptico. Una pregunta natural en análisis numérico para éstos problemas sería : ¿ La solución aproximada posee la propiedad del principio del máximo? . Éstos problemas tienen la propiedad que satisfacen el principio del máximo discreto (DMP), o también denominado principio de monotonía del método numérico.

En modelación matemática de varios fenómenos físicos, el Principio del Máximo Débil refleja la no negatividad natural de cantidades físicas como la temperatura, concentración, densidad, etc. La validez del Principio del Máximo Débil a nivel discreto es por lo tanto crítico para la confiabilidad de modelos numéricos. Aproximaciones con concentración negativa o flujos de calor que van de lugares más fríos a lugares más cálido son definitivamente no deseables y considerados no confiables.

El Principio del Máximo Débil ha sido estudiado y analizado por muchas décadas. Los primeros resultados que tratan con el método de diferencias finitas aparecen en 1964, donde se prueba que existe la matriz inversa y es positiva (ver [1]). Otro trabajo donde abordan un problema elíptico en forma de no divergencia con el Principio del Máximo Débil se encuentra en [2], donde se estudia la convergencia de ciertas técnicas iterativas comunes para la solución de los sistemas lineales resultantes.

Los esquemas monótonos existentes se derivan principalmente por una diferencia finita o un enfoque de volumen finito. Un inconveniente de éstos enfoques es que a menudo no está claro cómo analizar teóricamente los esquemas derivados de ésta manera. Así, nos motivaron a buscar esquemas monótonos que caen dentro del marco variacional de elementos finitos estándar, y por lo tanto, su análisis teórico es más directo. Una condición suficiente conocida para que un esquema sea monótono es que la correspondiente matriz de rigidez sea una M -matriz.

Un esquema monótono lineal usualmente tiene solo una precisión de primer orden, lo cual es un inconveniente bastante indeseable, y ciertamente limita su utilidad en los cálculos computacionales. Sin embargo, los esquemas monótonos lineales siguen siendo significativos de muchas maneras, por ejemplo, como una herramienta para diseñar técnicas iterativas y de preconditionamiento eficientes para resolver otros esquemas más sofisticados tales como.

Un sistema lineal con una M -matriz a partir de problemas con convección dominada se puede resolver de manera eficiente, por ejemplo, mediante el método de Gauss-Seidel, y la convergencia de la iteración de Gauss-Seidel se puede acelerar dramáticamente con un ordenamiento apropiado de incógnitas.

En este trabajo se analiza la existencia y unicidad de las soluciones de la ecuación de convección-difusión-reacción del tipo estacionario, se presentará el esquema discreto de Galerkin y se detallará el esquema EAFEM (ver [5]), el cual tiene varias características interesantes.

El esquema EAFEM es un esquema de elementos finitos con una formulación variacional estándar (pero con una forma bilineal modificada) por medio de las funciones lineales a trozos habituales para los espacios discretos, su derivación es completamente diferente de los otros enfoques conocidos, y no utiliza (explícitamente) las técnicas de upwinding (como verificar las direcciones de flujo).

Este esquema EAFEM fue parcialmente motivado por el trabajo de Markowich y Zlamal [7] y Brezzi, Marini y Pietra [8] y presentado por Xu Zikatanov en [5].

Contenido

Introducción	7
1. Preliminares	13
1.1. Teorema de Lax–Milgram	14
1.2. Espacios de Sobolev y Operador Trazas	22
1.3. Identidades de Green y Principio del Máximo Débil	29
1.4. Esquema de Galerkin	33
1.5. Esquema Monótono de Elementos Finitos	36
2. Problema Continuo Elíptico	39
2.1. Problema Modelo	39
2.2. Existencia y unicidad de la solución continua	40

3. Existencia y unicidad de las soluciones discretas	47
3.1. Esquema discreto de Galerkin	47
3.2. Esquema Monótono del Borde Promediado (EAFEM)	53
3.3. Ejemplos numéricos	70
3.3.1. Ejemplo 1	70
4. Conclusiones	85
Conclusiones	86
5. Sugerencias	87
Sugerencias	87
Referencias Bibliográficas	89

Capítulo 1

Preliminares

En éste primer capítulo se presentan algunas notaciones generales, se introducen definiciones y resultados generales sobre el Teorema de Lax–Milgram, el cual será una herramienta principal para el análisis de existencia y unicidad de la solución de la ecuación de convección–difusión–reacción del tipo estacionario a nivel continuo. También, se definen los espacios de Sobolev, Operadores de Trazas, Identidades de Green (fórmulas de integración por partes) y el Principio del Máximo Débil. Por último, se introduce la teoría para hacer el análisis de las soluciones discretas para la ecuación diferencial parcial elíptica de convección–difusión–reacción del tipo, presentando notaciones y definiciones generales sobre el Esquema Discreto de Galerkin (llamado también Esquema Clásico de Elementos Finitos) y el Esquema Monótono del Borde Promediado de Elementos Finitos.

Con respecto a los operadores diferenciales utilizados para una función escalar

$$w(\mathbf{x}) := w(x_1, x_2, \dots, x_d),$$

se tiene que $\partial_i w$ representará la derivada parcial de w con respecto a la variable x_i .

Además, los operadores gradiente y laplaciano para una función escalar w se definen como:

$$\nabla w := (\partial_1 w, \partial_2 w, \dots, \partial_d w)^t \in \mathbb{R}^d,$$

y

$$\Delta w := \sum_{i=1}^d \partial_{ii} w \in \mathbb{R},$$

respectivamente. Mientras que para una función vectorial, $\mathbf{w} = (w_1(x), w_2(x), \dots, w_d(x))$, se definen, respectivamente, los operadores divergencia y gradiente como:

$$\mathbf{div} \mathbf{w} := \nabla \cdot \mathbf{w} = \sum_{i=1}^d \partial_i w_i \in \mathbb{R},$$

y

$$\nabla \mathbf{w} := \begin{pmatrix} \frac{\partial w_1}{\partial x_1} & \frac{\partial w_1}{\partial x_2} & \dots & \frac{\partial w_1}{\partial x_d} \\ \frac{\partial w_2}{\partial x_1} & \frac{\partial w_2}{\partial x_2} & \dots & \frac{\partial w_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_d}{\partial x_1} & \frac{\partial w_d}{\partial x_2} & \dots & \frac{\partial w_d}{\partial x_d} \end{pmatrix}.$$

1.1. Teorema de Lax–Milgram

En esta sección se presentan conceptos básicos para los espacios de Hilbert, forma bilineal continua, la elipticidad de una forma (también llamada coersividad fuerte de una forma), y se presenta también el Teorema de Lax–Milgram y algunos resultados inmediatos de dicho teorema, los cuales se utilizarán para demostrar la existencia y unicidad de la solución del problema continuo.

DEFINICIÓN 1.1. Sea X un conjunto no vacío. Una **métrica** sobre el conjunto X es una función $d : X \times X \rightarrow \mathbb{R}$, que cumple las siguientes propiedades:

- a.- $d(x, y) \geq 0$, para todo $x, y \in X$.
- b.- $d(x, y) = 0$ si, y solo si, $x = y$.
- c.- $d(x, y) = d(y, x)$, para todo $x, y \in X$.
- d.- $d(x, z) \leq d(x, y) + d(y, z)$, para todo $x, y, z \in X$.

Observación 1.1. Al conjunto X provisto de una métrica d se le denomina **espacio métrico** (X, d) .

DEFINICIÓN 1.2. Sea X un espacio vectorial, se dice que la aplicación $\langle \cdot, \cdot \rangle_X : X \times X \rightarrow \mathbb{K}$ ($\mathbb{K} := \mathbb{R}$ o $\mathbb{K} := \mathbb{C}$) es un **producto interior** sobre el cuerpo \mathbb{K} si satisface:

- a.- $\langle x, x \rangle_X \geq 0$, para todo $x \in X$.
- b.- $\langle x, x \rangle_X = 0$ si, y sólo si, $x = 0$.
- c.- $\langle x, y \rangle_X = \overline{\langle y, x \rangle_X}$, para todo $x, y \in X$.
- d.- $\langle \alpha x + \beta y, z \rangle_X = \alpha \langle x, z \rangle_X + \beta \langle y, z \rangle_X$, para todo $\alpha, \beta \in \mathbb{K}$ y para todo $x, y, z \in X$.
- e.- $\langle x, \alpha y + \beta z \rangle_X = \overline{\alpha} \langle x, y \rangle_X + \overline{\beta} \langle x, z \rangle_X$, para todo $\alpha, \beta \in \mathbb{K}$ y para todo $x, y, z \in X$.

Observación 1.2. Al conjunto X provisto de un producto interior $\langle \cdot, \cdot \rangle_X$, se le denomina **espacio de Pre-Hilbert** $(X, \langle \cdot, \cdot \rangle_X)$.

DEFINICIÓN 1.3. Sea X un espacio vectorial sobre un cuerpo \mathbb{K} . Una función $\| \cdot \|_X : X \rightarrow \mathbb{K}$ es una **norma** sobre X , si para todo $x, y \in X$ y para todo $\alpha \in \mathbb{K}$, se satisfacen las siguientes propiedades:

- a.- $\|x\|_X \geq 0$, $\forall x \in X$.
- b.- $\|x\|_X = 0$ si, y sólo si, $x = 0$.
- c.- $\|\alpha x\|_X = |\alpha| \|x\|_X$, $\forall \alpha \in \mathbb{K}$ y $\forall x \in X$.
- d.- $\|x + y\|_X \leq \|x\|_X + \|y\|_X$, $\forall x, y \in X$.

TEOREMA 1.1 (Desigualdad de Cauchy–Schwarz). *Sea X un espacio vectorial, los vectores $x, y \in X$ satisfacen la siguiente desigualdad:*

$$|\langle x, y \rangle_X|^2 \leq \langle x, x \rangle_X \langle y, y \rangle_X.$$

Demostración. Sean $u, v \in X$ y $\alpha, \beta \in \mathbb{R}$. Luego, teniendo en cuenta las propiedades de producto interior, se tiene:

$$\begin{aligned} \langle \alpha u - \beta v, \alpha u - \beta v \rangle_X &\geq 0 \\ \langle \alpha u, \alpha u - \beta v \rangle_X - \langle \beta v, \alpha u - \beta v \rangle_X &\geq 0 \\ \langle \alpha u, \alpha u \rangle_X - \langle \alpha u, \beta v \rangle_X - \langle \beta v, \alpha u \rangle_X + \langle \beta v, \beta v \rangle_X &\geq 0 \\ \alpha \langle u, \alpha u \rangle_X - \alpha \langle u, \beta v \rangle_X - \beta \langle v, \alpha u \rangle_X + \beta \langle v, \beta v \rangle_X &\geq 0 \\ \alpha \bar{\alpha} \langle u, u \rangle_X - \alpha \bar{\beta} \langle u, v \rangle_X - \beta \bar{\alpha} \langle v, u \rangle_X + \beta \bar{\beta} \langle v, v \rangle_X &\geq 0 \\ |\alpha|^2 \langle u, u \rangle_X - \alpha \bar{\beta} \langle u, v \rangle_X - \beta \bar{\alpha} \langle v, u \rangle_X + |\beta|^2 \langle v, v \rangle_X &\geq 0. \end{aligned}$$

Luego, haciendo $\alpha = \langle u, v \rangle_X$ y $\beta = \langle u, u \rangle_X$ resulta,

$$\begin{aligned} |\langle u, v \rangle_X|^2 \langle u, u \rangle_X - \langle u, v \rangle_X \overline{\langle u, u \rangle_X} \langle u, v \rangle_X - \langle u, u \rangle_X \overline{\langle u, v \rangle_X} \langle v, u \rangle_X + |\langle u, u \rangle_X|^2 \langle v, v \rangle_X &\geq 0 \\ |\langle u, v \rangle_X|^2 \langle u, u \rangle_X - \langle u, v \rangle_X^2 \langle u, u \rangle_X - \langle u, u \rangle_X \langle v, u \rangle_X^2 + |\langle u, u \rangle_X|^2 \langle v, v \rangle_X &\geq 0 \\ |\langle u, v \rangle_X|^2 \langle u, u \rangle_X - 2|\langle u, v \rangle_X|^2 \langle u, u \rangle_X + |\langle u, u \rangle_X|^2 \langle v, v \rangle_X &\geq 0. \end{aligned}$$

De forma equivalente, se puede escribir,

$$\begin{aligned} |\langle u, v \rangle_X|^2 \langle u, u \rangle_X + |\langle u, u \rangle_X|^2 \langle v, v \rangle_X &\geq 2|\langle u, v \rangle_X|^2 \langle u, u \rangle_X \\ |\langle u, v \rangle_X|^2 + \langle u, u \rangle_X \langle v, v \rangle_X &\geq 2|\langle u, v \rangle_X|^2. \end{aligned}$$

Finalmente, se obtiene la siguiente desigualdad,

$$\langle u, u \rangle_X \langle v, v \rangle_X \geq |\langle u, v \rangle_X|^2.$$

□

DEFINICIÓN 1.4. *Sea $(X, \langle \cdot, \cdot \rangle_X)$ un espacio vectorial normado. Se dice **Espacio de Hilbert**, si X es completo con la norma inducida por el producto interior, es decir:*

$$\|x\|_X := \langle x, x \rangle_X^{\frac{1}{2}}, \quad \forall x \in X.$$

En el caso particular, de que $(X, \|\cdot\|_X)$ sea un espacio de Banach, se puede demostrar la siguiente igualdad

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \quad \forall x, y \in X,$$

conocida como **Ley del Paralelogramo**, identidad que permite comprobar que la norma proviene de un producto interior.

DEFINICIÓN 1.5. Sea $(X, \langle \cdot, \cdot \rangle_X)$ un espacio de Hilbert y sea $z \in X$ fijo. Entonces, para cada $x \in X$, se tiene que $\langle z, x \rangle \in \mathbb{R}$, es decir, se puede escribir la aplicación

$$\begin{aligned} \langle z, \cdot \rangle_X : X &\rightarrow \mathbb{R} \\ x &\mapsto \langle z, x \rangle_X. \end{aligned} \tag{1.1}$$

DEFINICIÓN 1.6. Sea $\widehat{F} : X \rightarrow \mathbb{R}$ un aplicación. Se dice que \widehat{F} es un **Funcional Lineal**, si se satisfacen las siguientes propiedades:

$$a.- \widehat{F}(x + y) = \widehat{F}(x) + \widehat{F}(y) \quad \forall x, y \in X.$$

$$b.- \widehat{F}(\alpha x) = \alpha \widehat{F}(x) \quad \forall \alpha \in \mathbb{R} \text{ y } \forall x \in X.$$

DEFINICIÓN 1.7. Dado $\widehat{F} : X \rightarrow \mathbb{R}$ un funcional lineal, se llama **funcional acotado**, si existe una constante $M > 0$, tal que:

$$|\widehat{F}(x)| \leq M\|x\|_X \quad \forall x \in X.$$

DEFINICIÓN 1.8. Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert. Se define $\mathcal{L}(H_1, H_2)$ como el conjunto de todos los operadores lineales y acotados que se definen en H_1 y toman valores en H_2 , es decir, $T : H_1 \rightarrow H_2$.

DEFINICIÓN 1.9. Sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert. Se define $\mathcal{L}(H)$ como el conjunto de todos los operadores lineales y acotados que se definen en H y toman valores en H , es decir, $T : H \rightarrow H$.

DEFINICIÓN 1.10. Sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert donde se define un operador lineal T y un producto interno $\langle x, y \rangle_H$. Entonces, T^* es llamado **operador adjunto** de T , si se cumple que:

$$\langle Tx, y \rangle_H = \langle x, T^*y \rangle_H.$$

DEFINICIÓN 1.11. Sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert. Se llama **Espacio Dual de H** , al conjunto formado por todos los funcionales lineales y acotados que se definen en H , el cual se denota por H' . Además, H' es dotada con la norma,

$$\| \cdot \|_{H'} : H' \rightarrow \mathbb{R}$$

$$\widehat{F} \mapsto \|\widehat{F}\|_{H'} := \sup_{x \neq 0 \in H} \frac{\widehat{F}(x)}{\|x\|}.$$

Las siguientes definiciones son fundamentales para enunciar el Teorema de Lax–Milgram.

DEFINICIÓN 1.12. Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert. Se dice que una forma $B : H_1 \times H_2 \rightarrow \mathbb{R}$ es **bilineal**, si ella es lineal en cada una de sus componentes, es decir:

$$a.- B(\alpha x + \beta y, z) = \alpha B(x, z) + \beta B(y, z) \quad \forall x, y \in H_1, \forall z \in H_2 \text{ y } \forall \alpha, \beta \in \mathbb{R}.$$

$$b.- B(x, \alpha y + \beta z) = \alpha B(x, y) + \beta B(x, z) \quad \forall x \in H_1, \forall y, z \in H_2 \text{ y } \forall \alpha, \beta \in \mathbb{R}.$$

DEFINICIÓN 1.13. Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert. Se dice que una forma bilineal $B : H_1 \times H_2 \rightarrow \mathbb{R}$ es acotada, si existe una constante $M > 0$, tal que:

$$|B(x, y)| \leq M \|x\|_1 \|y\|_2 \quad \forall x \in H_1, \text{ y } \forall y \in H_2.$$

DEFINICIÓN 1.14. Sean $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert y $B : H \times H \rightarrow \mathbb{R}$ una forma bilineal. Se dice que B es H –**elíptica** (o fuertemente coersiva), si existe una constante $\alpha > 0$ tal que:

$$B(x, x) \geq \alpha \|x\|_H^2 \quad \forall x \in H.$$

TEOREMA 1.2 (Teorema de Lax–Milgram). Sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert y sea $B : H \times H \rightarrow \mathbb{R}$ una forma bilineal, acotada y H –elíptica, con constantes de acotamiento y elipticidad, dadas por M y α , respectivamente. Entonces, para cada $\widehat{F} \in H'$, existe un único $v \in H$, tal que

$$B(v, w) = \widehat{F}(w) \quad \forall w \in H, \tag{1.2}$$

y

$$\|v\| \leq \frac{1}{\alpha} \|\widehat{F}\|_{H'}.$$

Demostración. Ver Teorema 1.1 en [3]. □

Los siguientes resultados se utilizarán para enunciar el Lema de Lax–Milgram.

TEOREMA 1.3 (Teorema de Representación de Riesz). *Sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert. Entonces, para todo $\widehat{F} \in H'$, existe un único $z \in H$, tal que:*

$$\widehat{F}(x) = \langle z, x \rangle_H \quad \forall x \in H.$$

Además, $\|\widehat{F}\|_{H'} = \|z\|_H$.

Observación 1.3. El Teorema de Representación de Riesz induce un operador

$$\begin{aligned} \mathcal{R} : H' &\rightarrow H \\ \widehat{F} &\mapsto \mathcal{R}(\widehat{F}) = z, \end{aligned}$$

el cual se conoce como operador de Riesz. Este operador está bien definido, dado que z es único, y que además es una isometría, es decir,

$$\|\mathcal{R}(\widehat{F})\|_H = \|z\|_H = \|\widehat{F}\|_{H'}.$$

Al elemento $z \in H$ se le llama **representante de Riesz**. A su vez, de acuerdo a lo anterior

$$\widehat{F}(x) = \langle \mathcal{R}(\widehat{F}), x \rangle_H \quad \forall x \in H.$$

DEFINICIÓN 1.15. *Dados $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert reales y una forma bilineal continua $B : H_1 \times H_2 \rightarrow \mathbb{R}$. Se define el funcional $\widehat{F}_v : H_2 \rightarrow \mathbb{R}$, por*

$$\widehat{F}_v(w) := B(v, w) \quad \forall w \in H_2. \tag{1.3}$$

Observación 1.4. Dado que B es bilineal, es claro que \widehat{F}_v es lineal. Además, del hecho que B es continuo, se infiere

$$|\widehat{F}_v(w)| \leq M\|v\|_1\|w\|_2 \quad \forall w \in H_2.$$

Esto garantiza que \widehat{F}_v es un funcional lineal y acotado de H_2 , y

$$\|\widehat{F}_v\| \leq M\|v\|_1 \quad \forall v \in H_1. \tag{1.4}$$

Del análisis anterior se induce la siguiente definición

DEFINICIÓN 1.16. Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert. Se define el operador $\mathcal{B} : H_1 \rightarrow H'_2$, como:

$$\mathcal{B}(v) := \widehat{F}_v, \quad \forall v \in H_1. \quad (1.5)$$

Además, de la linealidad de B en la primera componente y la desigualdad (1.4), es claro que \mathcal{B} es lineal y acotado con

$$\|\mathcal{B}\|_{\mathcal{L}(H_1, H'_2)} \leq M.$$

DEFINICIÓN 1.17 (Operador inducido). Dados $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert reales y $B : H_1 \times H_2 \rightarrow \mathbb{R}$ una forma bilineal continua. Se define el operador inducido como $\mathbb{B} : H_1 \rightarrow H_2$ tal que:

$$\mathbb{B} := \mathcal{R}_2 \circ \mathcal{B},$$

donde $\mathcal{R}_2 : H'_2 \rightarrow H_2$ denota la aplicación de Riesz. A su vez, de la linealidad y el acotamiento de \mathcal{R}_2 y \mathcal{B} se tiene:

$$\langle \mathbb{B}(v), w \rangle_2 = \langle \mathcal{R}_2(\mathcal{B}(v)), w \rangle_2 = \mathcal{B}(v)(w) \quad \forall (v, w) \in H_1 \times H_2,$$

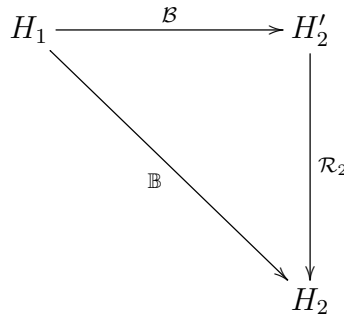
De (1.5) y (1.3) resulta:

$$\mathcal{B}(v)(w) = \widehat{F}_v(w) = B(v, w) \quad \forall (v, w) \in H_1 \times H_2,$$

lo que implica,

$$\langle \mathbb{B}(v), w \rangle_2 = B(v, w) \quad \forall (v, w) \in H_1 \times H_2.$$

o gráficamente



Para el siguiente resultado, es de interés definir $u \in H_1$ tal que:

$$\mathbb{B}(u) = \mathcal{R}_2(\widehat{F}_v).$$

Una condición suficiente y necesaria para que (1.2) tenga una única solución en cada $\widehat{F} \in H'_2$, es que el operador inducido por la forma bilineal sea biyectivo. Entonces, la biyectividad del operador inducido puede ser reformulado teniendo en cuenta el siguiente resultado.

LEMA 1.1 (Lema de Lax–Milgram). Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert con normas inducidas $\| \cdot \|_1$ y $\| \cdot \|_2$, respectivamente, y sea $\mathbb{B} \in \mathcal{L}(H_1, H_2)$. Entonces:

a.- \mathbb{B} es sobreyectivo si, y sólo sí, \mathbb{B}^* (operador inducido adjunto) es inyectivo y tiene rango cerrado, es decir, existe $\alpha > 0$, tal que:

$$\|\mathbb{B}^*(w)\|_1 \geq \alpha \|w\|_2 \quad \forall w \in H_2.$$

b.- \mathbb{B} es inyectivo, si y sólo sí,

$$\sup_{w \in H_2} \langle \mathbb{B}(v), w \rangle_2 > 0 \quad \forall v \in H_1, v \neq 0.$$

c.- \mathbb{B}^* es sobreyectivo si, y sólo sí, \mathbb{B} es inyectivo y tiene rango cerrado, es decir, existe $\alpha > 0$, tal que:

$$\|\mathbb{B}(v)\|_2 \geq \alpha \|v\|_1 \quad \forall v \in H_1.$$

d.- \mathbb{B}^* es inyectivo si, y sólo sí,

$$\sup_{v \in H_1} \langle \mathbb{B}(v), w \rangle_2 > 0, \quad \forall w \in H_2, w \neq 0$$

e.- \mathbb{B} es biyectivo si, y sólo sí, \mathbb{B}^* es biyectivo.

Demostración. Ver Lema 1.2 en [3].

□

LEMA 1.2 (Lema de Lax–Milgram Generalizado). Sean $(H_1, \langle \cdot, \cdot \rangle_1)$ y $(H_2, \langle \cdot, \cdot \rangle_2)$ espacios de Hilbert con normas inducidas $\| \cdot \|_1$ y $\| \cdot \|_2$, respectivamente, y sea $B : H_1 \times H_2 \rightarrow \mathbb{R}$ una forma bilineal continua. Suponga que:

(i) (Sobreyectividad) Existe $C > 0$, tal que:

$$\sup_{w \in H_2} \frac{B(v, w)}{\|w\|_2} \geq C\|v\|_1 \quad \forall v \in H_1, w \neq \theta.$$

(ii) (Inyectividad)

$$\sup_{v \in H_1} B(v, w) > 0 \quad \forall w \in H_2, w \neq \theta.$$

Entonces, para cada $\widehat{F} \in H_2'$, existe un único $v \in H_1$, tal que:

$$B(v, w) = \widehat{F}(w) \quad \forall w \in H_2 \quad \text{y} \quad \|v\|_1 \leq C^{-1} \|\widehat{F}\|_{H_2'}.$$

Además, las condiciones (i) y (ii) son también necesarias.

Demostración. Ver Teorema 1.2 en [3]. □

1.2. Espacios de Sobolev y Operador Trazas

En ésta sección se define el espacio de las funciones cuadrado integrables, el espacio de las funciones esencialmente acotadas, derivada en el sentido débil de una función, los espacios de Sóbolev, operadores de Trazas, y se presentan también resultados importantes de los operadores de Trazas para resolver las estimaciones *a priori* de las ecuaciones de convección–difusión–reacción de tipo estacionario.

Para empezar, considere $\Omega \subset \mathbb{R}^d$, con $d \in \mathbb{N}$, un dominio abierto y acotado de \mathbb{R}^d con frontera “suave” $\partial\Omega$, también denominada frontera poligonal o frontera Lipschitz continua, y considere $p \in [1, \infty]$ se denota el espacio de las funciones p integrables sobre el dominio Ω como $L^p(\Omega)$; mientras que el espacio de las funciones esencialmente acotadas sobre Ω , es denotada como $L^\infty(\Omega)$, es decir,

$$L^p(\Omega) := \left\{ v : \Omega \rightarrow \mathbb{R} \text{ tal que } \int_{\Omega} |v(x)|^p dx < \infty \right\},$$

y

$$L^\infty(\Omega) := \left\{ v : \Omega \rightarrow \mathbb{R} \text{ tal que } \text{ess sup}_{x \in \Omega} |v(x)| < \infty \right\},$$

donde, ess sup es el supremo esencial de v que se define como el ínfimo de las cotas superiores esenciales de v . Además, las normas de éstos espacios están definidas, respectivamente, de la siguiente forma:

$$\|v\|_{L^p(\Omega)} := \left(\int_{\Omega} |v(x)|^p dx \right)^{\frac{1}{p}},$$

y

$$\|v\|_{\infty, \Omega} := \text{ess sup}_{x \in \Omega} |v(x)|.$$

Para el caso $p = 2$, se escribirá $\|\cdot\|_{0, \Omega} = \|\cdot\|_{L^2(\Omega)}$ y se le denominará espacio de las funciones cuadrado integrables. Luego, se define el producto interno de $L^2(\Omega)$, como

$$\langle \cdot, \cdot \rangle_{0, \Omega} : L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}$$

$$(v, w) \mapsto \langle v, w \rangle_{0, \Omega} := \langle v, w \rangle_{L^2(\Omega)} = \int_{\Omega} v w \quad \forall v, w \in L^2(\Omega).$$

Observación 1.5. Se puede verificar fácilmente que $\|v\|_{0, \Omega}^2 = \langle v, v \rangle_{0, \Omega}$, es decir,

$$\|v\|_{0, \Omega}^2 = \int_{\Omega} |v|^2 = \int_{\Omega} v v = \langle v, v \rangle_{0, \Omega} \quad \forall v \in L^2(\Omega).$$

Entonces, la dupla $(L^2(\Omega), \langle \cdot, \cdot \rangle_{0, \Omega})$ es un espacio de Hilbert.

Ahora, se presentan dos desigualdades que serán de gran utilidad en el análisis de las estimaciones *a priori* para la ecuación de convección–difusión–reacción del tipo estacionario.

DEFINICIÓN 1.18 (Desigualdad de Cauchy–Schwarz para espacios cuadrado integrables). Si $v, w \in L^2(\Omega)$, entonces se satisface lo siguiente:

$$|\langle v, w \rangle_{0, \Omega}| \leq \left(\int_{\Omega} |v|^2 \right)^{\frac{1}{2}} \left(\int_{\Omega} |w|^2 \right)^{\frac{1}{2}} \quad \forall v, w \in L^2(\Omega). \quad (1.6)$$

DEFINICIÓN 1.19 (Desigualdad de Young para espacio cuadrado integrables). Si $v, w \in L^2(\Omega)$, entonces, existe $\xi > 0$ tal que :

$$\|v\|_{0, \Omega} \|w\|_{0, \Omega} \leq \frac{\|v\|_{0, \Omega}^2}{2\xi} + \frac{\xi \|w\|_{0, \Omega}^2}{2} \quad \forall v, w \in L^2(\Omega). \quad (1.7)$$

Para definir los espacios de Sobolev, se presentarán algunas definiciones previas sobre funciones con soporte compacto y derivada de una función en el sentido débil.

DEFINICIÓN 1.20 (Soporte Compacto). Sea $C(\Omega)$ el conjunto de las funciones continuas sobre Ω . Entonces, dado $\varphi \in C(\Omega)$, se define el soporte de Ω como el conjunto

$$\text{sop}(\varphi) = \overline{\{x \in \Omega : \varphi(x) \neq 0\}}.$$

DEFINICIÓN 1.21. Sea $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, y sea $C^\infty(\Omega)$ el conjunto de funciones infinitamente derivables en el sentido clásico (con límite). A su vez, sea $C(\Omega)$ el conjunto de las funciones continuas sobre Ω . Se define el conjunto de funciones $C^\infty(\Omega)$ con soporte compacto, como:

$$C_0^\infty(\Omega) := \{\varphi \in C^\infty(\Omega) : \text{sop}(\varphi) \text{ es compacto en } \Omega\}.$$

DEFINICIÓN 1.22. El espacio vectorial $C_0^\infty(\Omega)$, provisto de una topología inducida por una familia de seminormas, se denota por $\mathcal{D}(\Omega)$ y se llama **Espacio de funciones test**.

DEFINICIÓN 1.23 (Multi-índice). Un multi-índice es una d -upla $\alpha := (\alpha_1, \dots, \alpha_d)$ con $\alpha_i \in \mathbb{N} \cup \{0\}$, $\forall i \in \{1, \dots, d\}$. La longitud de α se define por:

$$|\alpha| := \sum_{i=1}^d \alpha_i.$$

Además, $D^\alpha v$ corresponde a la notación multi-índice para las **derivadas débiles** de v

$$D^\alpha v := \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}, \quad y \quad D^{(0, \dots, 0)} v := v.$$

EJEMPLO 1.1. A continuación, se presentan algunos ejemplos de multi-índices para los espacios \mathbb{R}^2 y \mathbb{R}^3 .

1.- Si $\alpha = (1, 1)$ entonces $\partial^\alpha v = \frac{\partial^2 v}{\partial x_1 \partial x_2}$.

2.- Si $\alpha = (0, 3)$ entonces $\partial^\alpha v = \frac{\partial^3 v}{\partial x_1^0 \partial x_2^3} = \frac{\partial^3 v}{\partial x_2^3}$.

3.- Si $\alpha = (1, 0, 2)$ entonces $\partial^\alpha v = \frac{\partial^3 v}{\partial x_1 \partial x_2^0 \partial x_3^2} = \frac{\partial^3 v}{\partial x_1 \partial x_3^2}$.

Comentario. Para una función $v \in L^1(\Omega)$, se dice que $w \in L^1(\Omega)$ es su α -ésima **derivada débil** si, y sólo si,

$$\int_{\Omega} v D^\alpha \varphi dx = (-1)^{|\alpha|} \int_{\Omega} w \varphi dx \quad \forall \varphi \in C_0^\infty(\Omega), \quad (1.8)$$

y haciendo una analogía a la notación clásica de derivada, se denota por $D^\alpha v := w$.

Comentario. Para una función $v \in L^2(\Omega)$, se dice que $\frac{\partial v}{\partial x_i} \in L^2(\Omega)$, si existe $z_i \in L^2(\Omega)$, tal que

$$\int_{\Omega} v \frac{\partial \varphi}{\partial x_i} = - \int_{\Omega} z_i \varphi, \quad \forall \varphi \in C_0^\infty(\Omega), \quad \forall i \in \{1, 2, \dots, d\}.$$

En tal caso, se escribirá $\frac{\partial v}{\partial x_i} = z_i$.

Luego de definir la derivada de una función en el sentido débil, se presentan ahora los espacios de Sobolev.

DEFINICIÓN 1.24 (Espacios de Sobolev). Sea $m \in \mathbb{N}$ y $p \in [1, \infty]$. Se define el **Espacio de Sobolev de orden m** , como:

$$W^{m,p}(\Omega) := \{v \in L^p(\Omega) : D^\alpha v \in L^p(\Omega) \quad \forall \alpha \in \mathbb{N}^d : |\alpha| \leq m\},$$

con $D^\alpha v$ que denota la derivada hasta el orden m de v , pero en el sentido débil.

Además, la norma del espacio de Sobolev se define de la siguiente manera:

$$\|v\|_{m,p,\Omega} = \begin{cases} \left[\sum_{|\alpha| \leq m} \|D^\alpha v\|_{L^p(\Omega)}^p \right]^{\frac{1}{p}}, & \text{si } 1 \leq p < \infty, \\ \sum_{|\alpha| \leq m} \text{ess sup}_{\Omega} |D^\alpha v|, & \text{si } p = \infty. \end{cases}$$

Mientras que la semi-norma, es dada por:

$$|v|_{m,p,\Omega} = \begin{cases} \left[\sum_{|\alpha|=m} \|D^\alpha v\|_{L^p(\Omega)}^p \right]^{\frac{1}{p}}, & \text{si } 1 \leq p < \infty, \\ \sum_{|\alpha|=m} \text{ess sup}_{\Omega} |D^\alpha v|, & \text{si } p = \infty. \end{cases}$$

Observación 1.6. Cuando $p = 2$, se escribirá el espacio de Sobolev $W^{m,2}(\Omega)$ como $H^m(\Omega)$, con

$$|\cdot|_{m,\Omega} := |\cdot|_{m,2,\Omega} \quad \text{y} \quad \|\cdot\|_{m,\Omega} := \|\cdot\|_{m,2,\Omega}.$$

Observación 1.7. En particular, para $m = 1$ y $p = 2$, se tiene el espacio $H^1(\Omega)$, donde la norma del espacio $H^1(\Omega)$ es denotada por $\|\cdot\|_{1,\Omega}$ y se define como:

$$\|u\|_{1,\Omega}^2 := \|u\|_{0,\Omega}^2 + |u|_{1,\Omega}^2,$$

con $|\cdot|_{1,\Omega}$ definido por:

$$|v|_{1,\Omega}^2 := \|\nabla v\|_{0,\Omega}^2 = \sum_{j=1}^d \left\| \frac{\partial v}{\partial x_j} \right\|_{0,\Omega}^2.$$

A su vez, sobre $H^1(\Omega)$, se define el siguiente producto interno,

$$\langle v, w \rangle_{H^1(\Omega)} := \int_{\Omega} vw + \int_{\Omega} \nabla v \cdot \nabla w \quad \forall v, w \in H^1(\Omega).$$

Note que,

$$\langle v, v \rangle_{H^1(\Omega)} = \int_{\Omega} |v|^2 + \int_{\Omega} |\nabla v|^2 = \|v\|_{0,\Omega}^2 + |v|_{1,\Omega}^2 = \|v\|_{1,\Omega}^2.$$

Así, la dupla $(H^1(\Omega), \langle \cdot, \cdot \rangle_{1,\Omega})$ es también un espacio de Hilbert.

Luego de definir el espacio de Sobolev $H^1(\Omega)$, se define el subespacio $H_0^1(\Omega)$.

DEFINICIÓN 1.25. Sea Ω un conjunto distinto del vacío, se define el espacio

$$\begin{aligned} H_0^1(\Omega) &:= \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{1,\Omega}} \\ &:= \{z \in H^1(\Omega) : \exists \{z_j\}_{j \in \mathbb{N}} \subseteq C_0^\infty(\Omega) \text{ tal que } \|z - z_j\|_{1,\Omega} \xrightarrow{j \rightarrow \infty} 0\} \\ &:= \{z \in H^1(\Omega) : z = 0 \text{ en } \partial\Omega\}. \end{aligned}$$

Observación. La notación $z = 0$ en $\partial\Omega$, indica que z se anula en la frontera de Ω . Sin embargo, se debe recordar que z no necesariamente está definida puntualmente. Por lo tanto, la restricción de z a la frontera no tiene sentido. Esta definición debe ser entendida en el sentido de trazas que se definirá más adelante.

Se define ahora el espacio $H_D^1(\Omega)$. Para ello, se denota por Γ a la frontera del dominio, $\partial\Omega$, provista de las siguientes propiedades:

a.- $\Gamma = \Gamma_D \cup \Gamma_N$,

b.- $\Gamma_D \cap \Gamma_N = \emptyset$,

donde Γ_D se denomina frontera de Dirichlet, y Γ_N frontera de Neumann, tal y como se muestra en la Figura 1.1¹, teniendo en cuenta que para la Figura 1.1, $\Gamma_D = \Gamma_1$ y $\Gamma_N = \Gamma_2$.

¹<https://images.app.goo.gl/YLuvBTUsNeo8aEmt8>

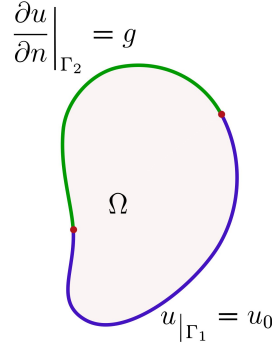


Figura 1.1: Dominio con frontera regular

DEFINICIÓN 1.26 (Espacio $H_D^1(\Omega)$). Sea $H^1(\Omega)$ el espacio de las funciones cuadrado integrables con primera derivada cuadrado integrable en el sentido débil. Entonces, se define el espacio $H_D^1(\Omega)$ como las funciones cuya traza es cero en la frontera Γ_D , es decir,

$$H_D^1(\Omega) := \{v \in H^1(\Omega) : v = 0 \text{ en } \Gamma_D\}.$$

Además, se denotará $[H_D^1(\Omega)]^{-1}$ al espacio dual de $H_D^1(\Omega)$ con respecto al producto interno de $L^2(\Omega)$. También, se denota $\langle \cdot, \cdot \rangle_{-1,1} := \langle \cdot, \cdot \rangle_{[H_D^1(\Omega)]^{-1}, H_D^1(\Omega)}$ como el producto dualidad entre $[H_D^1(\Omega)]^{-1}$ y $H_D^1(\Omega)$, es decir,

$$\begin{aligned} \langle \cdot, \cdot \rangle_{-1,1} : [H_D^1(\Omega)]^{-1} \times H_D^1(\Omega) &\rightarrow \mathbb{R} \\ (\widehat{F}, v) &\mapsto \langle \widehat{F}, v \rangle_{-1,1} = \int_{\Omega} \widehat{F}v. \end{aligned}$$

DEFINICIÓN 1.27 (Producto interno entre dos campos tensoriales). Dados los campos tensoriales $\tau := (\tau_{i,j})$ y $\zeta := (\zeta_{i,j})$, con $i, j = \{1, \dots, d\}$. Se define el producto interno como:

$$\tau : \zeta := \sum_{i,j=1}^d \tau_{i,j} \zeta_{i,j}.$$

A continuación, se define el operador traza. Para ello, se introducen las siguientes definiciones sobre distribuciones.

DEFINICIÓN 1.28. Sea $\{\varphi_k\}_{k \in \mathbb{N}}$ una sucesión de $\mathcal{D}(\Omega)$. Se dice que $\{\varphi_k\}_{k \in \mathbb{N}}$ **converge** a la función nula de $\mathcal{D}(\Omega)$, si existe un compacto $K_0 \subseteq \Omega$ tal que $\text{sop}(\varphi) \subseteq K_0$ para todo $k \in \mathbb{N}$, y

para toda d -upla α , esto es,

$$\sup_{x \in K_0} |\partial^\alpha \varphi_k(x)| \xrightarrow{k \rightarrow \infty} 0.$$

DEFINICIÓN 1.29. Una aplicación $v : \mathcal{D}(\Omega) \rightarrow \mathbb{C}$ se dice una *forma lineal* si:

$$v(\alpha\varphi + \beta\psi) = \alpha v(\varphi) + \beta v(\psi) \quad \forall \alpha, \beta \in \mathbb{C}, \forall \varphi, \psi \in \mathcal{D}(\Omega).$$

DEFINICIÓN 1.30. Una forma lineal $v : \mathcal{D}(\Omega) \rightarrow \mathbb{C}$ se dice una **distribución**, si ella es continua en $\mathcal{D}(\Omega)$, o equivalentemente, si para cada $\{\varphi_k\}_{k \in \mathbb{N}} \subseteq \mathcal{D}(\Omega)$ que converge a la función nula, se tiene que:

$$v(\varphi_k) \xrightarrow{k \rightarrow \infty} 0.$$

El espacio de distribuciones sobre Ω es denotado por $\mathcal{D}'(\Omega)$. Además, dado $v \in \mathcal{D}'(\Omega)$ y $\varphi \in \mathcal{D}(\Omega)$ se adopta la notación:

$$v(\varphi) = \langle v, \varphi \rangle.$$

En lo que sigue, se denotará por $\mathcal{D}(\overline{\Omega})$ (o $[C_0^\infty(\overline{\Omega})]$), al espacio de funciones de $\mathcal{D}(\overline{\Omega}')$, restringidas a Ω , sobre algún Ω' ; el cual contiene a Ω , es decir, $\Omega \subseteq \Omega'$.

DEFINICIÓN 1.31 (Operador de traza (ver [3])). Sea Ω un dominio acotado de \mathbb{R}^d , con $d \in \mathbb{N}$, con frontera “suave” Γ . Entonces, se define $\gamma_0 : \mathcal{D}(\overline{\Omega}) \rightarrow L^2(\Gamma)$ como:

$$\gamma_0(\varphi) = \varphi \Big|_{\Gamma} \quad \forall \varphi \in \mathcal{D}(\overline{\Omega}).$$

Además, existe $C > 0$ tal que:

$$\|\gamma_0(\varphi)\|_{0,\Gamma} \leq c \|\varphi\|_{1,\Omega} \quad \forall \varphi \in \mathcal{D}(\overline{\Omega}). \quad (1.9)$$

TEOREMA 1.4 (Teorema de Trazas en $H^1(\Omega)$). Sea Ω un abierto acotado de \mathbb{R}^d , $d \in \mathbb{N}$, con frontera poligonal Γ . Entonces, la aplicación $\gamma_0 : \mathcal{D}(\overline{\Omega}) \rightarrow L^2(\Gamma)$ se extiende por continuidad y densidad a una aplicación lineal y acotada $\gamma_0 : H^1(\Omega) \rightarrow L^2(\Gamma)$ tal que $\gamma_0(\varphi) = \varphi \Big|_{\Gamma}$ para todo $\varphi \in \mathcal{D}(\overline{\Omega})$.

Demostración. Ver [3]. □

1.3. Identidades de Green y Principio del Máximo Débil

Las identidades de Green son un conjunto de igualdades en cálculo vectorial, nombradas así en honor del matemático George Green, el mismo que descubrió el teorema de Green. Éstas identidades son necesarias para realizar la llamada integración por partes.

En lo que sigue, se define $\Omega \subset \mathbb{R}^d$, con $d \in \mathbb{N}$, un dominio abierto y acotado de \mathbb{R}^d con frontera poligonal, y se denota por $\mathbf{n} = (n_1, n_2, \dots, n_d)^t$, al vector normal unitario exterior al dominio Ω sobre la frontera $\partial\Omega$.

TEOREMA 1.5 (Fórmula de Integración por Partes). *Dados $u, v \in C^1(\Omega) \cap C(\overline{\Omega})$, entonces se satisface lo siguiente:*

$$\int_{\Omega} \left(u \frac{\partial v}{\partial x_i} + v \frac{\partial u}{\partial x_i} \right) = \int_{\partial\Omega} u v n_i \quad \forall i \in \{1, \dots, d\}. \quad (1.10)$$

Demostración. Sea un campo vectorial $\vec{G} = (G_1, G_2, \dots, G_d)^t \in [C^1(\Omega) \cap C(\overline{\Omega})]^d$. De acuerdo al Teorema de Gauss se sabe que:

$$\int_{\Omega} \mathbf{div} \vec{G} = \int_{\partial\Omega} \vec{G} \cdot \vec{n}. \quad (1.11)$$

En particular, dados $v, w \in C^1(\Omega) \cap C(\overline{\Omega})$ dos funciones escalares, elegimos

$$\vec{G} = (0, 0, \dots, 0, \underbrace{uv}_{i\text{-ésima}}, 0, \dots, 0)^t,$$

entonces, de (1.11), se llega a

$$\int_{\Omega} \frac{\partial(uv)}{\partial x_i} = \int_{\partial\Omega} u v n_i.$$

Así, obtenemos lo siguiente:

$$\int_{\Omega} \left(u \frac{\partial v}{\partial x_i} + v \frac{\partial u}{\partial x_i} \right) = \int_{\partial\Omega} u v n_i \quad \forall i \in \{1, \dots, d\}. \quad (1.12)$$

□

COROLARIO 1.1 (Primera Identidad de Green). *Sean $v \in C^1(\Omega) \cap C(\overline{\Omega})$ y $w \in C^2(\Omega) \cap C^1(\overline{\Omega})$, entonces se tiene:*

$$\int_{\Omega} \nabla w \cdot \nabla v + \int_{\Omega} v \Delta w = \int_{\partial\Omega} v \nabla w \cdot \mathbf{n}. \quad (1.13)$$

Demostración. Reemplazando ahora u por $\frac{\partial w}{\partial x_i}$ en (1.10), con $w \in C^2(\Omega) \cap C^1(\overline{\Omega})$, se sigue

$$\int_{\Omega} \left(\frac{\partial w}{\partial x_i} \frac{\partial v}{\partial x_i} + v \frac{\partial^2 w}{\partial x_i^2} \right) = \int_{\partial\Omega} \frac{\partial w}{\partial x_i} v n_i \quad \forall i \in \{1, \dots, d\},$$

y sumando sobre $i = 1, \dots, d$, se obtiene la siguiente igualdad,

$$\int_{\Omega} \sum_{i=1}^d \frac{\partial w}{\partial x_i} \frac{\partial v}{\partial x_i} + \int_{\Omega} v \sum_{i=1}^d \frac{\partial^2 w}{\partial x_i^2} = \int_{\partial\Omega} v \sum_{i=1}^d \frac{\partial w}{\partial x_i} n_i \quad \forall i \in \{1, \dots, d\}.$$

Por lo tanto, se concluye que:

$$\int_{\Omega} \nabla w \cdot \nabla v + \int_{\Omega} v \Delta w = \int_{\partial\Omega} v \nabla w \cdot \mathbf{n}.$$

□

Ahora, teniendo en cuenta la densidad de $\mathcal{D}(\overline{\Omega})$ en $H^1(\Omega)$ y el Teorema de Trazas en $H^1(\Omega)$, se deriva el siguiente resultado.

TEOREMA 1.6 (Fórmula de Integración por Partes). *Sea Ω un dominio acotado de \mathbb{R}^d , $d \in \mathbb{N}$, con frontera Lipschitz continua en $\partial\Omega$. Entonces, para cada $v, w \in H^1(\Omega)$ se tiene:*

$$\int_{\Omega} v \frac{\partial w}{\partial x_i} = - \int_{\Omega} w \frac{\partial v}{\partial x_i} + \int_{\partial\Omega} \gamma_0(v) \gamma_0(w) n_i \quad \forall i \in \{1, \dots, d\},$$

donde n_i es la i -ésima componente del vector normal.

Demostración. Ver Teorema 1.6 en [3].

□

Del teorema anterior surge el siguiente resultado.

COROLARIO 1.2 (Identidades de Green). *Sea Ω un dominio acotado de \mathbb{R}^d , $d \in \mathbb{N}$, con frontera Lipschitz continua $\partial\Omega$. Entonces, para cada $v \in H^1(\Omega)$ y $w \in H^2(\Omega)$, se cumple:*

$$\int_{\Omega} v \Delta w = - \int_{\Omega} \nabla w \nabla v + \int_{\partial\Omega} \gamma_0(v) \gamma_0(\nabla w) \cdot \mathbf{n}.$$

Demostración. Ver Corolario 1.2 en [3].

□

EJEMPLO 1.2. A continuación se presentan dos ejemplos para el espacio 1 – dimensional.

1.- Sea $\Omega = (-1, 1)$. Calcular $v'(x)$, sabiendo que:

$$v(x) = |x| = \begin{cases} -x & , \quad -1 < x \leq 0 \\ x & , \quad 0 \leq x < 1 \end{cases}.$$

Note que v no es una función derivable en el sentido clásico para $x = 0$. Sin embargo, es posible calcular su derivada débil. En efecto, dado $\varphi \in C_0^\infty(]-1, 1[)$, se cumple lo siguiente:

$$\begin{aligned} \int_{-1}^1 v \varphi' &= \int_{-1}^0 v \varphi' + \int_0^1 v \varphi' \\ &= \int_{-1}^0 (-x) \varphi' + \int_0^1 (x) \varphi' = - \lim_{a \rightarrow -1^+} \int_a^0 x \varphi' + \lim_{b \rightarrow 1^-} \int_0^b x \varphi' \\ &= - \lim_{a \rightarrow -1^+} \left[- \int_a^0 x' \varphi + \left(x \varphi \right)_a^0 \right] + \lim_{b \rightarrow 1^-} \left[- \int_0^b x' \varphi + \left(x \varphi \right)_0^b \right] \\ &= - \lim_{a \rightarrow -1^+} \left[- \int_a^0 (1) \varphi + \left((0) \varphi(0) - (a) \varphi(a) \right) \right] \\ &\quad + \lim_{b \rightarrow 1^-} \left[- \int_0^b (1) \varphi + \left((b) \varphi(b) - (0) \varphi(0) \right) \right] \\ &= \int_{-1}^0 \varphi - \left((0) \varphi(0) - (-1) \varphi(-1) \right) - \int_0^1 \varphi + \left((1) \varphi(1) - (0) \varphi(0) \right). \end{aligned}$$

Recordemos que φ evaluado en los extremos del intervalo es nulo, es decir $\varphi(1) = \varphi(-1) = 0$, entonces,

$$\int_{-1}^1 v \varphi' = \int_{-1}^0 \varphi - \int_0^1 \varphi = - \int_{-1}^1 z \varphi.$$

Por lo tanto,

$$v'(x) = z(x) = \begin{cases} -1 & , \quad -1 < x \leq 0 \\ 1 & , \quad 0 \leq x < 1 \end{cases}.$$

2.- Sea $\Omega = (0, 1)$. Calcular $v'(x)$, sabiendo que:

$$v(x) = \begin{cases} 2x & , \quad x \in (0, \frac{1}{2}) \\ 2 - 2x & , \quad x \in (\frac{1}{2}, 1) \end{cases},$$

En efecto dado $\varphi \in C_0^\infty([0, 1])$, se sigue que:

$$\begin{aligned}
 \int_0^1 v\varphi' &= \int_0^{\frac{1}{2}} v\varphi' + \int_{\frac{1}{2}}^1 v\varphi' \\
 &= - \int_0^{\frac{1}{2}} v'\varphi + \left(v\varphi\right)_0^{\frac{1}{2}} - \int_{\frac{1}{2}}^1 v'\varphi + \left(v\varphi\right)_{\frac{1}{2}}^1 \\
 &= - \int_0^{\frac{1}{2}} 2\varphi + \left(2\left(\frac{1}{2}\right)\varphi\left(\frac{1}{2}\right) - 2(0)\varphi(0)\right) \\
 &\quad - \int_{\frac{1}{2}}^1 (-2)\varphi + \left(\left[2 - 2\left(\frac{1}{2}\right)\right]\varphi\left(\frac{1}{2}\right) - [2 - 2(0)]\varphi(0)\right) \\
 &= - \int_0^{\frac{1}{2}} 2\varphi - \int_{\frac{1}{2}}^1 (-2)\varphi.
 \end{aligned}$$

Entonces:

$$v'(x) = z(x) = \begin{cases} 2 & , \quad 0 < x < \frac{1}{2} \\ -2 & , \quad \frac{1}{2} < x < 1 \end{cases}.$$

En lo que sigue, se presenta El Principio del Máximo Débil, el cual es una característica importante de las ecuaciones elípticas de segundo orden, que las distingue de las ecuaciones de orden superior y de los sistemas de ecuaciones. Además, de sus muchas aplicaciones, el Principio del Máximo Débil proporciona estimaciones puntuales que conducen a una teoría más definitiva.

Para presentar el Teorema del Principio del Máximo Débil, se considera el siguiente operador:

$$\widehat{\mathbf{L}}v := \nabla \cdot (\tilde{\alpha}\nabla v + \tilde{\mathbf{b}}v) + \tilde{c}\nabla u + \tilde{\gamma}v,$$

donde $\tilde{\alpha}$, $\tilde{\mathbf{b}}$, \tilde{c} y $\tilde{\gamma}$ son funciones medibles en un dominio Ω , con $0 < \tilde{\alpha}_0 \leq \tilde{\alpha} \leq \tilde{\alpha}_1$, $\tilde{\gamma}_0 > |\tilde{\gamma}|$ y $\|\tilde{\mathbf{b}}\|_{\infty, \Omega} < c$.

Además, asumiendo una de las siguientes desigualdades,

$$\text{i.- } \int_{\Omega} (\tilde{\gamma}v - \tilde{\mathbf{b}}\nabla u) \leq 0, \quad \forall v \in C_0^1(\Omega) \quad \forall v \geq 0.$$

$$\text{ii.- } \int_{\Omega} (\tilde{\gamma}v + \tilde{c}\nabla v) \leq 0 \quad \forall v \in C_0^1(\Omega), \quad \forall v \geq 0.$$

$$\text{iii.- } \begin{bmatrix} \tilde{\alpha} & \tilde{\beta} \\ -\tilde{c} & -\tilde{\gamma} \end{bmatrix} \geq 0,$$

se obtiene el siguiente resultado.

TEOREMA 1.7 (Principio del Máximo Débil). *Sea $v \in H^1(\Omega)$ tal que $\hat{\mathbf{L}}v \geq 0$ (o bien $\hat{\mathbf{L}}v \leq 0$) en Ω . Entonces,*

$$\sup_{\Omega} v \geq \sup_{\partial\Omega} v^+ \quad (\text{o } \inf_{\Omega} v \leq \inf_{\partial\Omega} v^-),$$

donde, $v^+ = \max\{v, 0\}$ y $v^- = \max\{-v, 0\}$.

Demostración. Ver Teorema 8.1 en [4]. □

COROLARIO 1.3 (Principio del Máximo Débil). *Sea $v \in H_D^1(\Omega)$ tal que $\hat{\mathbf{L}}v = 0$ en Ω . Entonces, $v = 0$ en Ω .*

Demostración. Ver Corolario 8.2 en [4]. □

1.4. Esquema de Galerkin

Los Esquemas de Elementos Finitos Clásicos, son un procedimiento numérico que permite obtener una aproximación numérica de la solución de una ecuación diferencial, ordinaria o parcial, bajo ciertas condiciones iniciales y de frontera, donde la idea principal es buscar funciones bases en un espacio de dimensión finita. Uno de éstos esquemas clásicos, es el llamado Esquema de Galerkin, este método convierte un problema de operador continuo (tal como una ecuación diferencial) en un problema discreto. El objetivo del método de Galerkin es la producción de un sistema lineal de ecuaciones construido por su forma de matriz que puede ser utilizado para calcular la solución mediante un programa informático.

Se define ahora el Esquema de Galerkin. Para ello, sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert, y sea $\hat{a} : H \times H \rightarrow \mathbb{R}$ una forma bilineal acotada y H -elíptica. Además, sea $\hat{F} : H \rightarrow \mathbb{R}$ un funcional lineal y acotado. Suponga que existe un único $v \in H$, tal que:

$$\hat{a}(v, w) = \hat{F}(w) \quad \forall w \in H.$$

La idea de éste esquema es tratar de encontrar una solución v_n , que pertenezca a un espacio de dimensión finita H_n (subespacio del espacio de dimensión infinita H), de tal forma que cada vez que $n \rightarrow \infty$, entonces $v_n \rightarrow v$.

Sea $\{H_n\}_{n \in \mathbb{N}}$ una sucesión de subespacios de H , de dimensión finita n y considere el problema:

$$\left\{ \begin{array}{l} \text{Hallar } v_n \in H_n \text{ tal que :} \\ \hat{a}(v_n, w_n) = \hat{F}(w_n) \quad \forall w_n \in H_n. \end{array} \right.$$

Note que \hat{F} restringido a H_n pertenece al dual de H_n , y $\hat{a} : H_n \times H_n \rightarrow \mathbb{R}$ es una forma bilineal acotada. Entonces, se tiene la misma estructura que en el caso continuo en H ; y por lo tanto, se puede aplicar el Teorema de Lax–Milgram o Lema de Lax–Milgram Generalizado para asegurar existencia y unicidad.

TEOREMA 1.8 (Teorema de Lax–Milgram Discreto). *Suponga que existe $\alpha_n > 0$ tal que $\hat{a}(w_n, w_n) \geq \alpha_n \|w_n\|_H^2 \quad \forall w_n \in H_n$. Entonces, existe una única $v_n \in H_n$, tal que:*

$$\hat{a}(v_n, w_n) = \hat{F}(w_n) \quad \forall w_n \in H_n.$$

Además,

$$\|v_n\|_H \leq \frac{1}{\alpha_n} \left\| \hat{F} \right\|_{H_n'} = \frac{1}{\alpha_n} \sup_{w_n \in H_n} \frac{|\hat{F}(w_n)|}{\|w_n\|_H} \quad \forall w_n \neq \vec{0}.$$

Demostración. Directo del Teorema de Lax–Milgram continuo. □

LEMA 1.3. (Lema de Lax–Milgram Generalizado Discreto) *Sea $\hat{a}_n : H_n \times H_n \rightarrow \mathbb{R}$ una forma bilineal acotada. Entonces, para todo $\hat{F}_n \in H_n'$, existe un único $v_n \in H_n$ tal que:*

$$\hat{a}_n(v_n, w_n) = \hat{F}_n(w_n) \quad \forall w_n \in H_n.$$

i.- Existe $\alpha_n > 0$, tal que:

$$\sup_{w_n \in H_n} \frac{\widehat{a}_n(w_n, v_n)}{\|w_n\|_H} \geq \alpha_n \|v_n\|_{H_n} \quad \forall w_n \neq \vec{0}.$$

ii.- O bien

$$\sup_{w_n \in H_n} \frac{\widehat{a}_n(v_n, w_n)}{\|w_n\|_H} \geq \alpha_n \|v_n\|_{H_n} \quad \forall w_n \neq \vec{0}.$$

En cualquier caso se obtiene que:

$$\|v_n\|_H \leq \frac{1}{\alpha_n} \|\widehat{F}_n\|_{H'_n}.$$

Demostración. Es una aplicación directa del Lema de Lax–Milgram Generalizado, teniendo en cuenta que en dimensión finita, inyectividad es equivalente a sobreyectividad. \square

Ahora, para el cálculo de la solución del Esquema de Galerkin, se procede de la siguiente forma.

Sea H_n un subespacio de H de dimensión finita n , y sea $\{e_1, \dots, e_n\}$ una base de H_n . Entonces, se tiene el siguiente problema discreto:

Hallar $v_n \in H_n$ tal que:

$$\widehat{a}(v_n, e_i) = \widehat{F}(e_i) \quad \forall i = \{1, \dots, n\}. \quad (1.14)$$

A su vez, como $v_n \in H_n$ (espacio de dimensión finita), existen $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ (incógnitas) tales que v_n se puede escribir como combinación lineal de las incógnitas por las funciones bases, es decir:

$$v_n = \sum_{j=1}^n \alpha_j e_j,$$

donde el problema discreto (1.14) se reescribe como:

Hallar $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tal que:

$$\sum_{j=1}^n \alpha_j \widehat{a}(e_j, e_i) = \widehat{F}(e_i) \quad \forall i = \{1, \dots, n\}.$$

Equivalentemente,

$$\underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\vec{\alpha}} \underbrace{\begin{bmatrix} \hat{a}(e_1, e_1) & \hat{a}(e_2, e_1) & \hat{a}(e_3, e_1) & \cdots & \hat{a}(e_n, e_1) \\ \hat{a}(e_1, e_2) & \hat{a}(e_2, e_2) & \hat{a}(e_3, e_2) & \cdots & \hat{a}(e_n, e_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{a}(e_1, e_n) & \hat{a}(e_2, e_n) & \hat{a}(e_3, e_n) & \cdots & \hat{a}(e_n, e_n) \end{bmatrix}}_A = \underbrace{\begin{bmatrix} \hat{F}(e_1) \\ \hat{F}(e_2) \\ \vdots \\ \hat{F}(e_n) \end{bmatrix}}_{\vec{f}}.$$

Matricialmente se escribe como: Hallar $\vec{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n$ tal que: $A\vec{\alpha} = \vec{f}$, donde

- $A \in \mathcal{M}_{n \times n}(\mathbb{C})$, llamada **matríz de rigidez** y $a_{ij} = a(e_j, e_i) \quad \forall i, j \in \{1, \dots, n\}$.
- $\vec{f} = (f_1, \dots, f_n)^t$ se llama **vector carga**, además, $f_j = \hat{F}(e_j) \quad \forall j = \{1, \dots, n\}$.

1.5. Esquema Monótono de Elementos Finitos

Los Esquemas monótonos de Elementos Finitos surgen debido a que los esquemas clásicos de elementos finitos no son suficientemente eficientes para calcular la solución numérica de ecuaciones diferenciales de convección–difusión–reacción con convección dominante. Éstos esquemas monótonos de elementos finitos se caracterizan por satisfacer el Principio del Máximo Débil, tanto a nivel continuo como a nivel discreto.

A continuación, se definen algunos conceptos previos para presentar el Esquema Monótono de Elementos Finitos.

DEFINICIÓN 1.32. Sea $\hat{L}v$ un operador elíptico. Se llamará **propiedad de monoticidad continua** o **propiedad de monotonía continua**, a la siguiente desigualdad,

$$\text{Si } (\hat{L}v)(x) \geq 0 \text{ para todo } x \in \Omega, \text{ entonces } v(x) \geq 0 \text{ para todo } x \in \Omega. \quad (1.15)$$

Ésta propiedad es independiente de $\frac{|\tilde{b}(x)|}{\tilde{\alpha}(x)}$.

Los casos que son interesantes estudiar para las aplicaciones es cuando la ecuación de convección–difusión–reacción, es de convección dominada, es decir, cuando $\frac{|\tilde{\mathbf{b}}(x)|}{\tilde{\alpha}(x)} \gg 1 \quad \forall x \in \Omega$.

DEFINICIÓN 1.33 (Propiedad de Monotonía Discreta). Sea $\hat{\mathbf{L}}_h$ la versión discreta del operador $\hat{\mathbf{L}}$, y si $V_h \subset V$ es el espacio de elementos finitos, entonces:

$$(\hat{\mathbf{L}}_h^{-1} f_h)(x) \geq 0 \text{ para todo } x \in \Omega, \text{ si } f_h^i \geq 0, \quad (1.16)$$

donde f_h^i es la función f evaluada en todas las funciones bases de V_h .

Un esquema monótono que satisface la condición (1.16), será llamado, a partir de ahora, como **esquema monótono de elementos finitos**.

DEFINICIÓN 1.34. (*M*–matriz) Sea $Z \in \mathcal{M}_{\mathbb{R} \times \mathbb{R}}(\mathbb{C})$ una matriz arbitraria. Se dice que Z es una *M*–matriz, si es irreducible y satisface lo siguiente:

$$a.- Z_{jj} > 0 \quad \forall j.$$

$$b.- Z_{ij} \leq 0 \quad \forall i, j : i \neq j.$$

$$c.- Z_{jj} \geq \sum_{i=1, i \neq j}^{N_h} |Z_{ij}| \quad \forall j.$$

$$d.- Z_{jj} \leq \sum_{i=1, i \neq j}^{N_h} |Z_{ij}| \text{ para algún } j.$$

Observación 1.8. Si la matriz de rigidez correspondiente a $\hat{\mathbf{L}}_h$ es una *M*–matriz, entonces se tiene (1.16), de ésta manera, se dice que el esquema es monótono de elementos finitos.

El siguiente teorema es de utilidad para demostrar las estimaciones *a priori* de los esquemas discretos de la ecuación de convección–difusión–reacción del tipo estacionario.

TEOREMA 1.9 (Bramble–Hilbert). Sea Ω un dominio acotado en \mathbb{R}^d , $d \in \mathbb{N}$, y asuma que \hat{F} es un funcional lineal en $W^{m,p}(\Omega)$ tal que,

$$a.- \hat{F}(v) \geq 0 \quad \forall v \in W^{m,p}(\Omega).$$

$$b.- \hat{F}(v + w) \leq \hat{F}(v) + \hat{F}(w) \quad \forall v, w \in W^{m,p}(\Omega).$$

$$c.- \widehat{F}(v) \leq C\|v\|_{m,p} \quad \forall v \in W^{m,p}(\Omega).$$

$$d.- \widehat{F}(v) = 0 \quad \forall v \in \mathbb{P}_{m-1}(\Omega).$$

Entonces, existe una constante $C = C(\Omega) > 0$ tal que $\widehat{F}(v) \leq C|v|_{m,p}$.

TEOREMA 1.10 (Desigualdad de Poincaré). *Sea p tal que $1 \leq p < \infty$, y sea Ω un subconjunto con al menos un borde acotado. Entonces, existe una constante $c > 0$, dependiendo sólo de Ω y de p , tal que, para cada función $v \in W_0^{1,p}(\Omega)$, se cumple:*

$$\|v\|_{L^p(\Omega)} \leq C\|\nabla v\|_{L^p(\Omega)}.$$

Observación 1.9. Si $W_0^{1,p}(\Omega) = H_0^1(\Omega)$, se tiene la siguiente desigualdad

$$|v|_{1,\Omega}^2 \geq c\|v\|_{1,\Omega}^2 \quad \forall v \in H_0^1(\Omega).$$

Capítulo 2

Problema Continuo Elíptico

En éste capítulo se presenta el problema modelo, el cual está caracterizado por una ecuación diferencial lineal elíptica en forma de divergencia, con condiciones iniciales de Neumann no homogéneas y condición de Dirichlet homogénea. Se analizará la existencia y unicidad de la solución continua de la ecuación de convección–difusión–reacción del tipo estacionario, utilizando como herramienta principal el Teorema de Lax–Milgram. Por último, se analizará la monotonicidad del esquema a nivel continuo.

2.1. Problema Modelo

Considere como problema modelo, a la ecuación diferencial parcial elíptica lineal de convección–difusión–reacción del tipo estacionario, en forma de divergencia con condiciones de contorno de Dirichlet y Neumann. Es decir:

$$\left\{ \begin{array}{ll} \mathbf{L}u := -\nabla \cdot (\varepsilon(x)\nabla u(x) + \mathbf{b}(x)u(x)) + \gamma(x)u(x) = f(x) & x \in \Omega, \\ u(x) = 0 & x \in \Gamma_D, \\ (\varepsilon(x)\nabla u(x) + \mathbf{b}(x)u(x)) \cdot \mathbf{n} = g & x \in \Gamma_N, \end{array} \right. \quad (2.1)$$

donde $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, es un dominio acotado con frontera poligonal $\partial\Omega = \Gamma_N \cup \Gamma_D$, y $\Gamma_N \cap \Gamma_D = \emptyset$ y se denota por \mathbf{n} al vector normal unitario exterior a $\partial\Omega$.

Para el problema (2.1) se consideran las siguientes hipótesis:

$$\mathcal{A}_1) \quad \varepsilon_1 \geq \varepsilon(x) \geq \varepsilon_0 > 0.$$

$$\mathcal{A}_2) \quad \gamma(x) > \gamma_0 > 0.$$

$$\mathcal{A}_3) \quad \varepsilon \in W^{1,\infty}(\Omega), \mathbf{b} \in [W^{1,\infty}(\Omega)]^d, \gamma \in W^{1,\infty}(\Omega).$$

$$\mathcal{A}_4) \quad f \in L^2(\Omega) \text{ y } g \in L^2(\Gamma_N).$$

Observación 2.1. En lo que sigue, se le llama problema variacional continuo al problema modelo.

2.2. Existencia y unicidad de la solución continua

Para demostrar existencia y unicidad del problema variacional continuo (2.1), se deriva primero a un problema equivalente llamado problema débil. Para ello, se multiplica la primera ecuación del problema (2.1) por una función test adecuada y utilizando la fórmula del corolario 1.2 de la Identidad de Green, se obtiene que para todo $v \in H_D^1(\Omega)$,

$$\begin{aligned} & - \int_{\Omega} \nabla(\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\Omega} \gamma uv = \int_{\Omega} f v \\ & - \left(- \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\partial\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} v \right) + \int_{\Omega} \gamma uv = \int_{\Omega} f v \\ & \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v - \int_{\partial\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} v + \int_{\Omega} \gamma uv = \int_{\Omega} f v \\ & \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v - \int_{\Gamma_D} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} v - \int_{\Gamma_N} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} v + \int_{\Omega} \gamma uv = \int_{\Omega} f v. \end{aligned}$$

Teniendo en cuenta que $u = 0$ en Γ_D , y que $(\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} = g$ en Γ_N , entonces:

$$\int_{\Gamma_D} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n} v = 0,$$

y

$$\int_{\Gamma_N} (\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n}v = \int_{\Gamma_N} gv,$$

obteniéndose,

$$\int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v - \int_{\Gamma_N} gv + \int_{\Omega} \gamma uv = \int_{\Omega} fv \quad \forall v \in H_D^1(\Omega).$$

De forma equivalente, se puede escribir:

$$\int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\Omega} \gamma uv = \int_{\Omega} fv + \int_{\Gamma_N} gv \quad \forall v \in H_D^1(\Omega).$$

Luego, se define la forma $a(\cdot, \cdot)$ como:

$$\begin{aligned} a : H_D^1(\Omega) \times H_D^1(\Omega) &\rightarrow \mathbb{R} \\ (u, v) &\mapsto \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\Omega} \gamma uv. \end{aligned}$$

y el funcional $F(\cdot)$ como:

$$\begin{aligned} F : H_D^1(\Omega) &\rightarrow \mathbb{R} \\ v &\mapsto \int_{\Omega} fv + \int_{\Gamma_N} gv. \end{aligned}$$

Por lo tanto, el problema variacional (2.1) se reescribe como un problema débil, dado por:

$$\left\{ \begin{array}{l} \text{Hallar } u \in H_D^1(\Omega) \text{ tal que:} \\ a(u, v) = F(v) \quad \forall v \in H_D^1(\Omega). \end{array} \right. \quad (2.2)$$

Demostrar existencia y unicidad de la solución del problema variacional continuo (2.1) es equivalente a demostrar existencia y unicidad de la solución del problema débil (2.2). Para ello, se verifican las hipótesis del Teorema de Lax–Milgram :

- a.- $a(\cdot, \cdot)$ es una forma bilineal, es decir $\forall s, t \in \mathbb{R}$ y $\forall u, v, w \in H_D^1(\Omega)$, se demuestra la linealidad en cada componente:

■

$$\begin{aligned}
& a(su + tv, w) \\
&= \int_{\Omega} (\varepsilon \nabla(su + tv) + \mathbf{b}(su + tv)) \cdot \nabla w + \int_{\Omega} \gamma(su + tv)w \\
&= \int_{\Omega} s(\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla w + \int_{\Omega} t(\varepsilon \nabla v + \mathbf{b}v) \cdot \nabla w + \int_{\Omega} s(\gamma uw) + t(\gamma vw) \\
&= s \left(\int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla w + \int_{\Omega} \gamma uw \right) + t \left(\int_{\Omega} (\varepsilon \nabla v + \mathbf{b}v) \cdot \nabla w + \int_{\Omega} \gamma vw \right) \\
&= sa(u, w) + ta(v, w).
\end{aligned}$$

■

$$\begin{aligned}
& a(u, sv + tw) \\
&= \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla(sv + tw) + \int_{\Omega} \gamma u(sv + tw) \\
&= \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot s \nabla v + \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot t \nabla w + \int_{\Omega} (s(\gamma uv) + t(\gamma uw)) \\
&= s \left(\int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\Omega} \gamma uv \right) + t \left(\int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla w + \int_{\Omega} \gamma uw \right) \\
&= sa(u, v) + ta(u, w).
\end{aligned}$$

b.- $a(\cdot, \cdot)$ es una forma acotada. En efecto, $\forall u, v \in H_D^1(\Omega)$, se tiene:

$$\begin{aligned}
|a(u, v)| &= \left| \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v + \int_{\Omega} \gamma uv \right| \\
&\leq \left| \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \cdot \nabla v \right| + \left| \int_{\Omega} \gamma uv \right|.
\end{aligned}$$

Luego, utilizando desigualdad (1.6) se deduce,

$$\begin{aligned}
|a(u, v)| &\leq \left(\int_{\Omega} |\varepsilon \nabla u + \mathbf{b}u|^2 \right)^{\frac{1}{2}} \left(\int_{\Omega} |\nabla v|^2 \right)^{\frac{1}{2}} + \left(\int_{\Omega} |\gamma u|^2 \right)^{\frac{1}{2}} \left(\int_{\Omega} |v|^2 \right)^{\frac{1}{2}} \\
&= \|\varepsilon \nabla u + \mathbf{b}u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|\gamma u\|_{0,\Omega} \|v\|_{0,\Omega} \\
&\leq \|\varepsilon \nabla u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|\mathbf{b}u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|\gamma u\|_{0,\Omega} \|v\|_{0,\Omega}.
\end{aligned} \tag{2.3}$$

Observe que:

■

$$\begin{aligned}
 \|\varepsilon \nabla u\|_{0,\Omega}^2 &= \int_{\Omega} |\varepsilon \nabla u|^2 = \int_{\Omega} |\varepsilon|^2 |\nabla u|^2 \\
 &\leq \max \{ |\varepsilon(x)|^2 : x \in \Omega \} \int_{\Omega} |\nabla u|^2 = \|\varepsilon\|_{\infty,\Omega}^2 \|\nabla u\|_{0,\Omega}^2 \\
 &\Rightarrow \|\varepsilon \nabla u\|_{0,\Omega} \leq \|\varepsilon\|_{\infty,\Omega} \|\nabla u\|_{0,\Omega}.
 \end{aligned} \tag{2.4}$$

■

$$\begin{aligned}
 \|\gamma u\|_{0,\Omega}^2 &= \int_{\Omega} |\gamma u|^2 \leq \int_{\Omega} |\gamma|^2 |u|^2 \\
 &\leq \max \{ |\gamma(x)|^2 : x \in \Omega \} \int_{\Omega} |u|^2 = \|\gamma\|_{\infty,\Omega}^2 \|u\|_{0,\Omega}^2 \\
 &\Rightarrow \|\gamma u\|_{0,\Omega} \leq \|\gamma\|_{\infty,\Omega} \|u\|_{0,\Omega}.
 \end{aligned} \tag{2.5}$$

Reemplazando (2.4) y (2.5) en (2.3), se obtiene:

$$|a(u, v)| \leq \|\varepsilon\|_{\infty,\Omega} \|\nabla u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|\mathbf{b}\|_{\infty,\Omega} \|u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|\gamma\|_{\infty,\Omega} \|u\|_{0,\Omega} \|v\|_{0,\Omega}.$$

Haciendo $\varepsilon_{\Omega} = \|\varepsilon\|_{\infty,\Omega}$, $\mathbf{b}_{\Omega} = \|\mathbf{b}\|_{\infty,\Omega}$ y $\gamma_{\Omega} = \|\gamma\|_{\infty,\Omega}$, se sigue que:

$$\begin{aligned}
 |a(u, v)| &\leq \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \} (\|\nabla u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|u\|_{0,\Omega} \|\nabla v\|_{0,\Omega} + \|u\|_{0,\Omega} \|v\|_{0,\Omega}) \\
 &\leq \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \} (\|\nabla u\|_{0,\Omega}^2 + \|u\|_{0,\Omega}^2 + \|u\|_{0,\Omega}^2)^{\frac{1}{2}} + (\|\nabla v\|_{0,\Omega}^2 + \|\nabla v\|_{0,\Omega}^2 + \|v\|_{0,\Omega}^2)^{\frac{1}{2}} \\
 &\leq \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \} (\|\nabla u\|_{0,\Omega}^2 + 2\|u\|_{0,\Omega}^2)^{\frac{1}{2}} + (2\|\nabla v\|_{0,\Omega}^2 + \|v\|_{0,\Omega}^2)^{\frac{1}{2}} \\
 &\leq 2 \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \} (\|\nabla u\|_{0,\Omega}^2 + \|u\|_{0,\Omega}^2)^{\frac{1}{2}} + (\|\nabla v\|_{0,\Omega}^2 + \|v\|_{0,\Omega}^2)^{\frac{1}{2}} \\
 &\leq 2 \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \} \|u\|_{1,\Omega} \|v\|_{1,\Omega}.
 \end{aligned}$$

Así, $a(\cdot, \cdot)$ es una forma continua con constante de acotamiento $M = 2 \max \{ \varepsilon_{\Omega}, \mathbf{b}_{\Omega}, \gamma_{\Omega} \}$.

c.- Ahora, se demuestra que $F(\cdot)$ es un funcional lineal. En efecto, dado $v \in V$, se tiene:

$$\begin{aligned}
 F(sv + tw) &= \int_{\Omega} f(sv + tw) + \int_{\Gamma_N} g(sv + tw) \\
 &= \int_{\Omega} sfv + \int_{\Omega} tfw + \int_{\Gamma_N} sgv + \int_{\Gamma_N} tgw \\
 &= s \left(\int_{\Omega} fv + \int_{\Gamma_N} gv \right) + t \left(\int_{\Omega} fw + \int_{\Gamma_N} gw \right) \\
 &= sF(v) + tF(w).
 \end{aligned}$$

d.- $F(\cdot)$ es un funcional acotado. En efecto, dado $v \in V$, se tiene:

$$|F(v)| = \left| \int_{\Omega} fv + \int_{\Gamma_N} gv \right| \leq \left| \int_{\Omega} fv \right| + \left| \int_{\Gamma_N} gv \right|.$$

De la desigualdad (1.6) resulta

$$\begin{aligned}
 |F(v)| &\leq \left(\int_{\Omega} |f|^2 \right)^{\frac{1}{2}} \left(\int_{\Omega} |v|^2 \right)^{\frac{1}{2}} + \left(\int_{\Gamma_N} |g|^2 \right)^{\frac{1}{2}} \left(\int_{\Gamma_N} |v|^2 \right)^{\frac{1}{2}} \\
 &= \|f\|_{0,\Omega} \|v\|_{0,\Omega} + \|g\|_{0,\Gamma_N} \|v\|_{0,\Gamma_N}.
 \end{aligned}$$

Y utilizando el Teorema de Desigualdad de Trazas se llega a:

$$|F(v)| \leq (\|f\|_{0,\Omega} + \|g\|_{0,\Gamma_N}) \|v\|_{0,\Omega}.$$

Así, $F(\cdot)$ es un funcional continuo con constante de acotamiento $M = \|f\|_{0,\Omega} + \|g\|_{0,\Gamma_N}$.

e.- Se demuestra que $a(\cdot, \cdot)$ es una forma H -elíptica. Para ello, sea $u \in V$, entonces

$$\begin{aligned}
 a(u, u) &= \int_{\Omega} (\varepsilon \nabla u + \mathbf{b}u) \nabla u + \int_{\Omega} \gamma |u|^2 \\
 &= \int_{\Omega} \varepsilon |\nabla u|^2 + \int_{\Omega} \mathbf{b}u \cdot \nabla u + \int_{\Omega} \gamma |u|^2.
 \end{aligned}$$

Utilizando las hipótesis \mathcal{A}_1 , \mathcal{A}_2 y propiedades fundamentales de integrales se deduce:

$$\begin{aligned}
 a(u, u) &\geq \varepsilon_0 \int_{\Omega} |\nabla u|^2 + \int_{\Omega} \mathbf{b}u \cdot \nabla u + \gamma_0 \int_{\Omega} |u|^2 \\
 &= \varepsilon_0 \|\nabla u\|_{0,\Omega}^2 + \int_{\Omega} \mathbf{b}u \cdot \nabla u + \gamma_0 \|u\|_{0,\Omega}^2 \\
 &\geq \varepsilon_0 \|\nabla u\|_{0,\Omega}^2 - \int_{\Omega} |\mathbf{b}u \cdot \nabla u| + \gamma_0 \|u\|_{0,\Omega}^2.
 \end{aligned}$$

De la desigualdad de Cauchy–Schwarz y de Young, respectivamente, resulta

$$\begin{aligned}
a(u, u) &\geq \varepsilon_0 \|\nabla u\|_{0,\Omega}^2 - \|\mathbf{b}\|_{\infty,\Omega} \|u\|_{0,\Omega} \|\nabla u\|_{0,\Omega} + \gamma_0 \|u\|_{0,\Omega}^2 \\
&\geq \varepsilon_0 \|\nabla u\|_{0,\Omega}^2 - \|\mathbf{b}\|_{\infty,\Omega} \left(\frac{\|u\|_{0,\Omega}^2}{2 \left(\frac{\varepsilon_0}{\|\mathbf{b}\|_{\infty,\Omega}} \right)} + \frac{\left(\frac{\varepsilon_0}{\|\mathbf{b}\|_{\infty,\Omega}} \right) \|\nabla u\|_{0,\Omega}^2}{2} \right) + \gamma_0 \|u\|_{0,\Omega}^2 \\
&\geq \varepsilon_0 \|\nabla u\|_{0,\Omega}^2 - \|\mathbf{b}\|_{\infty,\Omega} \left(\frac{\|\mathbf{b}\|_{\infty,\Omega} \|u\|_{0,\Omega}^2}{2\varepsilon_0} + \frac{\varepsilon_0 \|\nabla u\|_{0,\Omega}^2}{2\|\mathbf{b}\|_{\infty,\Omega}} \right) + \gamma_0 \|u\|_{0,\Omega}^2 \\
&= \left(\varepsilon_0 - \frac{\varepsilon_0}{2} \right) \|\nabla u\|_{0,\Omega}^2 + \left(\gamma_0 - \frac{\|\mathbf{b}\|_{\infty,\Omega}^2}{2\varepsilon_0} \right) \|u\|_{0,\Omega}^2 \\
&= \frac{\varepsilon_0}{2} \|\nabla u\|_{0,\Omega}^2 + \left(\gamma_0 - \frac{\|\mathbf{b}\|_{\infty,\Omega}^2}{2\varepsilon_0} \right) \|u\|_{0,\Omega}^2 \\
&\geq \min \left\{ \frac{\varepsilon_0}{2}, \gamma_0 - \frac{\|\mathbf{b}\|_{\infty,\Omega}^2}{2\varepsilon_0} \right\} (\|\nabla u\|_{0,\Omega}^2 + \|u\|_{0,\Omega}^2) \\
&= \min \left\{ \frac{\varepsilon_0}{2}, \gamma_0 - \frac{\|\mathbf{b}\|_{\infty,\Omega}^2}{2\varepsilon_0} \right\} \|u\|_{1,\Omega}^2.
\end{aligned}$$

Considerando $\gamma_0 > \frac{\|\mathbf{b}\|_{\infty,\Omega}^2}{2\varepsilon_0}$ se tiene que $a(\cdot, \cdot)$ es H –elíptica.

Así, surge el siguiente resultado que nos garantiza que el problema continuo está bien puesto.

TEOREMA 2.1. *Asumiendo las hipótesis $\mathcal{A}_1 - \mathcal{A}_4$ se tiene que existe una única solución del problema débil (2.2).*

Demostración. Dado que $a(\cdot, \cdot)$ es una forma bilineal, continua y H –elíptica, y que el funcional $F(\cdot)$ es lineal y acotado, se sigue del Teorema de Lax–Milgram, que existe una única solución para el problema (2.2). \square

Observación 2.2. Note que las constantes del operador en estudio \mathbf{L} son funciones medibles que satisfacen $0 < \varepsilon_0 \leq \varepsilon(x) \leq \varepsilon_1$, $0 < \gamma_0 < \gamma(x)$, $\|\mathbf{b}\|_{\infty,\Omega}$. Además, se cumple lo siguiente

$$\begin{vmatrix} \tilde{\alpha} & \tilde{\beta} \\ -\tilde{c} & -\tilde{\gamma} \end{vmatrix} = \begin{vmatrix} -\varepsilon & -\beta \\ 0 & -\gamma \end{vmatrix} = (-\varepsilon(x))(-\gamma(x)) > 0.$$

Lo que implica que \mathbf{L} satisface las condiciones del Teorema del Principio del Máximo Débil, es decir el operador \mathbf{L} satisface la propiedad de monotonicidad continua:

$$\text{Si } (\mathbf{L}u)(x) \geq 0 \text{ para todo } x \in \Omega \text{ entonces } u(x) \geq 0 \text{ para todo } x \in \Omega. \quad (2.6)$$

Capítulo 3

Existencia y unicidad de las soluciones discretas

En éste capítulo se presenta el esquema discreto de Galerkin y el esquema EAFEM para el problema (2.1), y para cada uno de ellos se analiza la existencia y unicidad de sus soluciones aproximadas utilizando los Lemas de Lax–Milgram. Por último, presentamos un ejemplo numérico, el cual servirá para comparar la eficiencia de los métodos.

3.1. Esquema discreto de Galerkin

En ésta sección se presenta una discretización del problema variacional continuo (2.2) utilizando el esquema de Galerkin clásico y polinomios lineales, continuos a trozos, el cual será llamado esquema discreto de Galerkin. Para ello, se considera una partición \mathcal{T}_h que pertenece a $\{\mathcal{T}_h\}_h$, con $h > 0$, una familia de particiones de Ω , que satisfacen las siguientes propiedades:

- 1.- Cualquier par de triángulos de \mathcal{T}_h son disjuntos.
- 2.- Cualquier par de triángulos comparten un lado en común, o tienen un vértice en común.

$$3.- \sup_{h>0} \sup_{K \in \mathcal{T}_h} \frac{h_K}{\rho_K} \leq \kappa, \text{ con la constante } \kappa \text{ independiente de la malla,}$$

donde, h_K es el diámetro del elemento K (dimensión del lado de mayor longitud) y ρ_K es el diámetro de la circunsferencia inscrita en el elemento K . Para tener una idea más clara sobre malla regular observe la Figura 3.1 ¹. Luego, se define el subespacio de elementos finitos V_h como:

$$V_h =: \{v_h \in C^0(\bar{\Omega}) \cap V : v_h|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h\} \subset V,$$

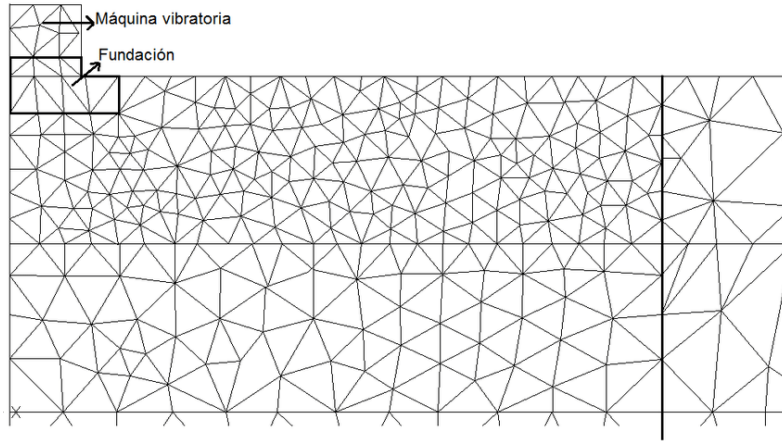


Figura 3.1: Malla regular

donde:

- \mathbb{P}_1 es el espacio de polinomios de grado 1.
- N_h es la cantidad de nodos de la malla.
- φ_i con $i \in \{1, \dots, N_h\}$ denotan las funciones bases nodales en V_h .
- V es un espacio de dimensión infinita.

¹<https://images.app.goo.gl/5n859qTcJERSeukq9>

- $x_i, i = \{1, \dots, N_h\}$ denotan los grados de libertad de la triangulación.

Además,

$$\varphi_i(x_j) = \begin{cases} \varphi_i(x_j) = 1, & i = j \\ \varphi_i(x_j) = 0, & j \neq i. \end{cases} \quad (3.1)$$

A continuación, se presenta el problema discreto de Galerkin para la ecuación de convección–difusión–reacción del tipo estacionario. Éste esquema de Galerkin se utiliza para estudiar y derivar un esquema monótono, denominado esquema monótono de elementos finitos del borde promediado (EAFEM) que se introduce en la siguiente sección.

$$\begin{cases} \text{Hallar } \tilde{u}_h \in V_h \text{ tal que} \\ a(\tilde{u}_h, v_h) = F(v_h) \quad \forall v_h \in V_h. \end{cases} \quad (3.2)$$

donde la forma bilineal $a(\cdot, \cdot)$ y el funcional lineal $F(\cdot)$ fueron definidos en la sección 2.1.

En lo que sigue, se presenta el operador de proyección ortogonal de Galerkin y algunos resultados que serán de utilidad para la demostración de la existencia y unicidad de la solución del esquema discreto.

DEFINICIÓN 3.1. Sea $P_h : V \rightarrow V_h$ la proyección ortogonal de Galerkin, tal que:

$$\tilde{a}(P_h v, w_h) = \tilde{a}(v, w_h) \quad \forall w_h \in V_h, \quad (3.3)$$

donde $\tilde{a}(v, w) := (\varepsilon \nabla v, \nabla w)$.

Se puede mostrar fácilmente que la forma $\tilde{a}(\cdot, \cdot)$ es una forma bilineal, continua y coersiva.

- Para demostrar la bilinealidad de la forma $a(\cdot, \cdot)$, demostraremos la linealidad en una sola componente, mientras que la linealidad de la otra componente se demuestra de forma análoga, por lo cual se omiten futuros detalles. Sean $r, s \in \mathbb{R}$ y $v, w, x \in V$.

$$\begin{aligned} \tilde{a}(rv + sw, x) &= \int_{\Omega} \varepsilon \nabla(rv + sw) \cdot \nabla x = r \int_{\Omega} \varepsilon \nabla v \cdot \nabla x + s \int_{\Omega} \varepsilon \nabla w \cdot \nabla x \\ &= r \tilde{a}(v, x) + s \tilde{a}(w, x). \end{aligned}$$

ii.- La forma bilineal $\tilde{a}(\cdot, \cdot)$ es continua, es decir para $v, w \in V$, se tiene:

$$\begin{aligned} |\tilde{a}(v, w)| &= \int_{\Omega} \varepsilon \nabla v \cdot \nabla w \leq \max \{ \varepsilon(x), x \in \Omega \} \int_{\Omega} \nabla v \cdot \nabla w \\ &= \varepsilon_1 \int_{\Omega} \nabla v \cdot \nabla w \leq \varepsilon_1 \|\nabla v\|_{0,\Omega} \|\nabla w\|_{0,\Omega} \\ &\leq \varepsilon_1 \|v\|_{1,\Omega} \|w\|_{1,\Omega}. \end{aligned}$$

iii.- La forma bilineal $\tilde{a}(\cdot, \cdot)$ es coersiva, es decir para $v \in V$, se cumple:

$$\begin{aligned} \tilde{a}(v, v) &= \int_{\Omega} \varepsilon \nabla v \cdot \nabla v \geq \min \{ \varepsilon(x), x \in \Omega \} \int_{\Omega} |\nabla v|^2 \\ &= \varepsilon_0 \|\nabla v\|_{0,\Omega}^2. \end{aligned}$$

Luego, aplicando el Lema de Poincaré se obtiene $\varepsilon_0 \|\nabla v\|_{0,\Omega}^2 \geq \varepsilon_0 \|v\|_{1,\Omega}^2$. Así,

$$\tilde{a}(v, v) \geq \varepsilon_0 \|v\|_{1,\Omega}^2.$$

LEMA 3.1. Dado $P_h : V \rightarrow V_h$ la proyección ortogonal de Galerkin. Entonces, existe una constante $C(\varepsilon)$ tal que:

$$\|P_h v\|_{1,\Omega} + h^{-1} \|v - P_h v\|_{0,\Omega} \leq C(\varepsilon) \|v\|_{1,\Omega} \quad \forall v \in V. \quad (3.4)$$

Demostración. Se prueba la desigualdad $\|P_h v\|_{1,\Omega} \leq C(\varepsilon) \|v\|_{1,\Omega}$. Para ello, hacemos $v = P_h(v)$ en la primera componente, y de la Definición 3.1 se obtiene,

$$\tilde{a}(P_h v, P_h v) = \tilde{a}(v, P_h v).$$

Utilizando la continuidad y la coersividad de la forma bilineal $\tilde{a}(\cdot, \cdot)$, resulta:

$$\varepsilon_0 \|P_h v\|_{1,\Omega}^2 \leq \tilde{a}(P_h v, P_h v) = \tilde{a}(v, P_h v) \leq \varepsilon_1 \|v\|_{1,\Omega} \|P_h v\|_{1,\Omega}.$$

De lo anterior se deduce que

$$\|P_h v\|_{1,\Omega} \leq \frac{\varepsilon_1}{\varepsilon_0} \|v\|_{1,\Omega}.$$

□

El siguiente resultado nos garantiza que existe una única solución para el problema discreto de Galerkin (3.2). Además la demostración del resultado se sigue de manera similar que el Lema 2.2 en [6].

LEMA 3.2. *Dado $h > 0$ suficientemente pequeño, existe $C > 0$ independiente de h tal que:*

$$\sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_{1,\Omega}} \geq C \|w_h\|_{1,\Omega} \quad \forall w_h \in V_h.$$

Demostración. Teniendo en cuenta que en espacios de dimensión finita la inyectividad implica sobreyectividad, demostraremos la inyectividad del operador inducido por la forma bilineal $a(\cdot, \cdot)$, es decir

$$\sup_{w_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{1,\Omega}} \geq C \|v_h\|_{1,\Omega} \quad \forall v_h \in V_h.$$

Aplicando el Lema de Lax–Milgram en (2.2), se tiene que el operador adjunto del operador inducido por la forma bilineal $a(\cdot, \cdot)$ es sobreyectivo, es decir,

$$\sup_{w \in V} \frac{a(w, v_h)}{\|w\|_{1,\Omega}} \geq C_0 \|v_h\|_{1,\Omega} \quad \forall v_h \in V_h, \quad (3.5)$$

Además, no es difícil ver que

$$a(P_h w, v_h) = a(w, v_h) - a(w - P_h(w), v_h), \quad (3.6)$$

luego, sumando $\tilde{a}(w - P_h(w), v_h) = 0$ en la ecuación (3.6) y de la Definición 3.1, resulta

$$\begin{aligned} a(P_h w, v_h) &= a(w, v_h) - a(w - P_h(w), v_h) + \tilde{a}(w - P_h(w), v_h) \\ &= a(w, v_h) + (\tilde{a} - a)(w - P_h(w), v_h). \end{aligned} \quad (3.7)$$

Observe que

$$\begin{aligned} (\tilde{a} - a)(w - P_h(w), v_h) &= \int_{\Omega} \varepsilon \nabla(w - P_h(w)) \cdot \nabla(v_h) - \int_{\Omega} \varepsilon \nabla(w - P_h(w)) \cdot \nabla(v_h) \\ &\quad - \int_{\Omega} \mathbf{b}(w - P_h(w)) \cdot \nabla(v_h) - \int_{\Omega} \gamma(w - P_h(w)) v_h \\ &= - \int_{\Omega} \mathbf{b}(w - P_h(w)) \cdot \nabla(v_h) - \int_{\Omega} \gamma(w - P_h(w)). \end{aligned} \quad (3.8)$$

Reemplazando ahora (3.8) en (3.7), se obtiene

$$a(P_h w, v_h) = a(w, v_h) - \int_{\Omega} \mathbf{b}(w - P_h(w)) \cdot \nabla(v_h) - \int_{\Omega} \gamma(w - P_h(w)).$$

Utilizando desigualdad de Cauchy–Schwarz y Young, se deduce:

$$\begin{aligned}
a(P_h w, v_h) &\geq a(w, v_h) - \|\mathbf{b}\|_{\infty, \Omega} \|w - P_h w\|_{0, \Omega} \|\nabla v_h\|_{0, \Omega} - \|\gamma\|_{\infty, \Omega} \|w - P_h w\|_{0, \Omega} \|v_h\|_{0, \Omega} \\
&\geq a(w, v_h) - (\|\mathbf{b}\|_{\infty, \Omega} \|\nabla v_h\|_{0, \Omega} + \|\gamma\|_{\infty, \Omega} \|v_h\|_{0, \Omega}) \|w - P_h w\|_{0, \Omega} \\
&\geq a(w, v_h) - (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} (\|\nabla v_h\|_{0, \Omega}^2 + \|v_h\|_{0, \Omega}^2)^{\frac{1}{2}} \|w - P_h w\|_{0, \Omega} \\
&\geq a(w, v_h) - (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} \|v_h\|_{1, \Omega} \|w - P_h w\|_{0, \Omega}.
\end{aligned}$$

Y del Lema 3.1 se cumple que:

$$a(P_h w, v_h) \geq a(w, v_h) - C(\varepsilon) (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega} \|w\|_{1, \Omega}.$$

Equivalentemente, la desigualdad anterior está dada por:

$$\frac{a(P_h w, v_h)}{\|P_h w\|_{1, \Omega}} \geq \frac{a(w, v_h)}{\|P_h w\|_{1, \Omega}} - \frac{C(\varepsilon) (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega} \|w\|_{1, \Omega}}{\|P_h w\|_{1, \Omega}}. \quad (3.9)$$

Por otra parte, gracias al Lema 3.1, se tiene que si,

$$\|P_h w\|_{1, \Omega} \leq \frac{\varepsilon_1}{\varepsilon_0} \|w\|_{1, \Omega}, \quad \text{entonces} \quad \frac{1}{\|P_h w\|_{1, \Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} \frac{1}{\|w\|_{1, \Omega}}.$$

Reemplazando la desigualdad anterior en (3.9), resulta

$$\frac{a(P_h w, v_h)}{\|P_h w\|_{1, \Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} \frac{a(w, v_h)}{\|w\|_{1, \Omega}} - C(\varepsilon) \frac{\varepsilon_0}{\varepsilon_1} (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega}. \quad (3.10)$$

Además,

$$\sup_{w \in V} \frac{a(P_h w, v_h)}{\|P_h w\|_{1, \Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} \sup_{w \in V} \frac{a(w, v_h)}{\|w\|_{1, \Omega}} - C(\varepsilon) \frac{\varepsilon_0}{\varepsilon_1} (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega}. \quad (3.11)$$

Usando el hecho de que

$$\sup_{w_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{1, \Omega}} \geq \sup_{w_h \in Im(P_h)} \frac{a(P_h w, v_h)}{\|P_h w\|_{1, \Omega}} = \sup_{w \in V} \frac{a(P_h w, v_h)}{\|P_h w\|_{1, \Omega}}. \quad (3.12)$$

y reemplazando (3.12) en (3.11) se llega a

$$\sup_{w_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{1, \Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} \sup_{w \in V} \frac{a(w, v_h)}{\|w\|_{1, \Omega}} - C(\varepsilon) \frac{\varepsilon_0}{\varepsilon_1} (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega}.$$

Así, de (3.5) se concluye

$$\sup_{w_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{1, \Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} c_0 \|v_h\|_{1, \Omega} - C(\varepsilon) (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h \|v_h\|_{1, \Omega} \quad (3.13)$$

$$\geq \frac{\varepsilon_0}{\varepsilon_1} (c_0 - C(\varepsilon) (\|\mathbf{b}\|_{\infty, \Omega}^2 + \|\gamma\|_{\infty, \Omega}^2)^{\frac{1}{2}} h) \|v_h\|_{1, \Omega}. \quad (3.14)$$

Finalmente, eligiendo h suficientemente pequeño se llega a

$$\sup_{w_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{1,\Omega}} \geq \frac{\varepsilon_0}{\varepsilon_1} c_0 \|v_h\|_{1,\Omega}.$$

□

3.2. Esquema Monótono del Borde Promediado (EAFEM)

En ésta sección se presenta el Esquema Monótono de Elementos Finitos del Borde Promediado, el cual se caracteriza por promediar adecuadamente los coeficientes de la ecuación diferencial parcial sobre las aristas de la triangulación. Éste esquema EAFEM es una aproximación del esquema clásico de Galerkin (3.2). Para ello, introduciremos la siguiente definición

Definición. (d -dimensional simplex) Sea $\{\bar{a}_j\}_{j=1}^{d+1}$, $d+1$ vectores de \mathbb{R}^d con $\bar{a}_j = (a_{ij})_{i=1}^d$, tales que la matriz

$$A = \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \cdots & \bar{a}_{d+1} \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,d+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{d,d+1} \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

es invertible. Entonces la cápsula convexa K generada por $\{\bar{a}_j\}_{j=1}^{d+1}$, se define por:

$$K = \left\{ x = \sum_{j=1}^{d+1} \alpha_j \bar{a}_j : 0 \leq \alpha_j \leq 1, \sum_{j=1}^{d+1} \alpha_j = 1 \right\}.$$

Los vectores $\{\bar{a}_j\}_{j=1}^{d+1}$ se llaman vértices del d -simplex K . Para tener una idea más clara de un conjunto d -dimensional simplex observe la Figura 3.2².

²<https://images.app.goo.gl/afRgC55ZNG3Ccfus7>

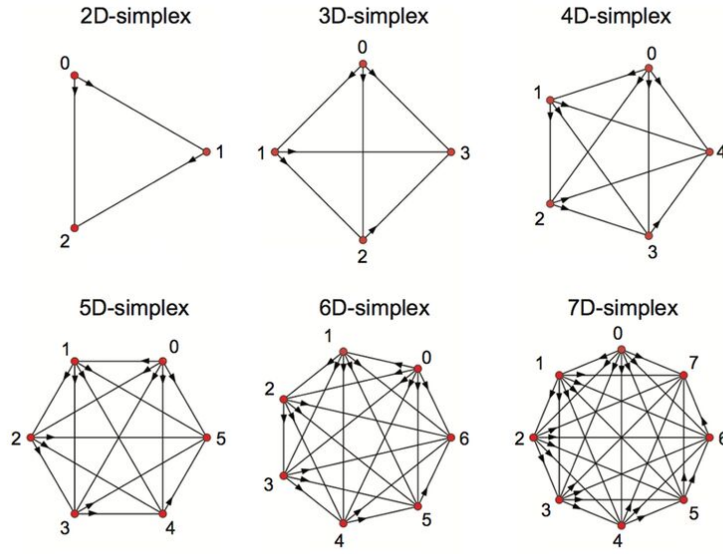


Figura 3.2: 1-7 dimensionales simplex

Observación 3.1. Notar que la invertibilidad de A garantiza que el sistema lineal

$$A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \\ \alpha_{d+1} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ 1. \end{pmatrix},$$

tiene una única solución para $\mathbf{x} \in \mathbb{R}^d$. Además

$$\sum_{j=1}^{d+1} \alpha_j \bar{a}_j = x, \quad \sum_{j=1}^{d+1} \alpha_j = 1.$$

Dado $K \in \mathcal{T}_h$ un elemento de la triangulación se presenta la siguiente notación;

- 1.- q_j ($1 \leq j \leq d+1$): el vértice de K ;
- 2.- $E_{i,j}$ o simplemente E : la arista que conecta los vértices q_i, q_j ;
- 3.- F_j : el $(d-1)$ -dimensional simplex opuesto al vértice q_j ;
- 4.- $\theta_{i,j}^K$ o θ_E^K : el ángulo entre F_i y F_j ;

5.- $\kappa_E^K: F_i \cap F_j$, el $(d-2)$ -dimensional simplex opuesto al borde E ;

6.- Para cualquier función continua ϕ en $E = E_{ij}$, se define:

$$\delta_E(\phi) := \phi(q_i) - \phi(q_j); \quad (3.15)$$

4.- $\mathbf{n}_E := q_i - q_j$ un vector direccional de E .

Se introduce la idea del método como en [5], empezando con el caso para la ecuación de Poisson, el cual detallaremos a continuación.

$$\begin{cases} -\Delta u = f, & x \in \Omega, \\ u = 0, & x \in \partial\Omega. \end{cases} \quad (3.16)$$

Teniendo en cuenta que $u_h, v_h \in V_h$ se pueden escribir como combinación lineal de escalares (incógnitas) que coinciden con las funciones u_h, v_h evaluada en las coordenadas del elemento K , se tiene:

$$\begin{aligned} \int_K \nabla u_h \cdot \nabla v_h dx &= \int_K \nabla \left(\sum_{i=1}^{d+1} u_h(q_i) \varphi_i \right) \cdot \nabla \left(\sum_{j=1}^{d+1} v_h(q_j) \varphi_j \right) dx \\ &= \sum_{i,j}^{d+1} u_h(q_i) v_h(q_j) \int_K \nabla \varphi_i \cdot \nabla \varphi_j = \sum_{i,j}^{d+1} a_{i,j}^K u_h(q_i) v_h(q_j). \end{aligned}$$

Denotando $a_{i,j}^K = \sum_{i,j}^{d+1} a_{i,j}^K u_h(q_i) v_h(q_j)$, el cual se llama matriz de rigidez, entonces se tiene,

$$\int_K \nabla u_h \cdot \nabla v_h dx = \sum_{i,j}^{d+1} a_{i,j}^K u_h(q_i) v_h(q_j).$$

Por otra parte, como $\sum_{i=1}^{d+1} \varphi_i = 1$, entonces $\sum_{i=1}^{d+1} \nabla \varphi_i = 0$. Así,

$$\sum_{i,j=1}^{d+1} \nabla \varphi_i \cdot \nabla \varphi_j = 0,$$

obteniéndose $a_{i,i}^K = - \sum_{i \neq j} a_{i,j}^K$, con lo que se infiere la siguiente identidad,

$$\int_K \nabla u_h \cdot \nabla v_h dx = - \sum_{i < j} a_{i,j}^K (u_h(q_i) - u_h(q_j))(v_h(q_i) - v_h(q_j)) \quad \forall u_h, v_h \in V_h. \quad (3.17)$$

Haciendo $\omega_E^K = -(a_{i,j}^K)$, con E conectando los vértices q_i, q_j se tiene,

$$-\sum_{i < j} a_{i,j}^K (u_h(q_i) - u_h(q_j))(v_h(q_i) - v_h(q_j)) = \sum_{E \subset K} \omega_E^K \delta_E u_h \delta_E v_h. \quad (3.18)$$

Reemplazando (3.18) en (3.17) resulta,

$$\int_K \nabla u_h \cdot \nabla v_h dx = \sum_{E \subset K} \omega_E^K \delta_E u_h \delta_E v_h.$$

Luego se suma sobre todos los triángulos de \mathcal{T}_h es decir,

$$\sum_{K \in \mathcal{T}_h} \int_K \nabla u_h \cdot \nabla v_h dx = \sum_{K \in \mathcal{T}_h} \sum_{E \subset K} \omega_E^K \delta_E u_h \delta_E v_h.$$

Lo que finalmente implica que,

$$\int_{\Omega} \nabla u_h \cdot \nabla v_h dx = \sum_{K \in \mathcal{T}_h} \sum_{E \subset K} \omega_E^K \delta_E u_h \delta_E v_h. \quad (3.19)$$

Donde el peso ω_E^K se define de la siguiente forma (ver [5]),

$$\omega_E^K = \frac{1}{d(d-1)} |\kappa_E^K| \cot \theta_E^K.$$

El siguiente lema dado en [5] nos garantiza que la matriz de rigidez de la ecuación de Poisson es M -matriz.

LEMA 3.3. *La matriz de rigidez para la ecuación de Poisson es una M -matriz si y sólo si para cualquier arista fija E la siguiente desigualdad sostiene,*

$$\omega_E \equiv \frac{1}{d(d-1)} \sum_{E \subset K} |\kappa_E^K| \cot \theta_E^K \geq 0, \quad (3.20)$$

donde $\sum_{E \subset K}$ significa la suma de todos los simplex K que contienen a E .

A continuación, se presenta la idea del método para el problema (2.1) y detallaremos el método siguiendo las ideas de [5].

En primer lugar consideramos el caso cuando $\mathbf{b} = \nabla c$, con $c \in C^1(\Omega)$ y denotamos por:

$$J(u_h) := \varepsilon \nabla u_h + \mathbf{b} u_h = \varepsilon \nabla u_h + u_h \nabla c.$$

Observe que,

$$J(u_h) = \varepsilon e^{\frac{-c}{\varepsilon}} \nabla(e^{\frac{c}{\varepsilon}} u_h) = \kappa \nabla(\phi_h),$$

donde, $\phi_h = e^\psi u_h$, $\kappa = \varepsilon e^{-\psi}$, $\psi = (c/\varepsilon)$. A su vez se observa lo siguiente,

$$\int_K J(u_h) \cdot \nabla v_h = \int_K (\kappa \nabla \phi_h) \cdot \nabla v_h. \quad (3.21)$$

Teniendo en cuenta que ϕ_h y v_h son funciones de un espacio de dimensión finita entonces,

$$\begin{aligned} \int_K (\kappa \nabla \phi_h) \cdot \nabla v_h &= \int_K \kappa \nabla \left(\sum_{i=1}^{N_h} \phi_h(q_i) \varphi_i \right) \cdot \nabla \left(\sum_{j=1}^{N_h} v_h(q_j) \varphi_j \right) \\ &= \int_K \kappa \left(\sum_{i=1}^{N_h} \phi_h(q_i) \nabla \varphi_i \right) \cdot \left(\sum_{j=1}^{N_h} v_h(q_j) \nabla \varphi_j \right) \\ &= \int_K \kappa \sum_{i,j=1}^{N_h} \phi_h(q_i) v_h(q_j) \nabla \varphi_i \cdot \nabla \varphi_j. \end{aligned}$$

Recuerde que $\phi(q_i)$ y $v_h(q_j)$, $\forall i, j = 1, \dots, N_h$ son constantes entonces,

$$\int_K (\kappa \nabla \phi_h) \cdot \nabla v_h = \sum_{i,j=1}^{N_h} \phi_h(q_i) v_h(q_j) \int_K \kappa (\nabla \varphi_i \cdot \nabla \varphi_j). \quad (3.22)$$

Reemplazando (3.22) en (3.21) resulta,

$$\int_K J(u_h) \cdot \nabla v_h = \sum_{i,j=1}^{N_h} \phi_h(q_i) v_h(q_j) \int_K \kappa (\nabla \varphi_i \cdot \nabla \varphi_j).$$

Para ese sistema se tiene que la correspondiente matriz de rigidez es,

$$\int_K \kappa (\nabla \varphi_i \cdot \nabla \varphi_j).$$

Procediendo de manera similar que en el caso de la Ecuación de Poisson se tiene,

$$\int_K J(u_h) \cdot \nabla v_h = - \sum_{i < j} \delta_E(\phi_h) \delta_E(v_h) \int_K \kappa (\nabla \varphi_i \cdot \nabla \varphi_j). \quad (3.23)$$

Luego dado que φ_i, φ_j , son funciones lineales, tenemos que,

$$\int_K \kappa \nabla \varphi_i \cdot \nabla \varphi_j = \left(\frac{1}{|K|} \int_K \kappa \right) \left(\int_K \nabla \varphi_i \cdot \nabla \varphi_j \right). \quad (3.24)$$

Recuerde que $\kappa = \varepsilon e^{-\psi}$ en (3.24),

$$\int_K \kappa \nabla \varphi_i \cdot \nabla \varphi_j = \left[\frac{1}{|K|} \int_K \varepsilon e^{-\psi} \right] \left[\int_K \nabla \varphi_i \cdot \nabla \varphi_j \right]. \quad (3.25)$$

Luego, se define el promedio armónico de $\varepsilon e^{-\psi_E}$ sobre E como:

$$\tilde{\varepsilon}_E(\mathbf{b}) := \left[\frac{1}{|\mathcal{T}_E|} \int_E \varepsilon^{-1} e^\psi ds \right]^{-1}, \quad (3.26)$$

el cual satisface lo siguiente,

$$\left[\frac{1}{|\mathcal{T}_E|} \int_E \varepsilon^{-1} e^\psi ds \right]^{-1} \approx \left[\frac{1}{|K|} \int_K \varepsilon e^{-\psi} \right]. \quad (3.27)$$

Así se obtiene la siguiente aproximación,

$$\int_K \kappa \nabla \varphi_i \cdot \nabla \varphi_j \approx \tilde{\varepsilon}_E(\mathbf{b}) \left[\int_K \nabla \varphi_i \cdot \nabla \varphi_j \right] = \tilde{\varepsilon}_E(\mathbf{b}) (a_{i,j}^K). \quad (3.28)$$

Reemplazando (3.28) en (3.23) se deduce,

$$\int_K J(u_h) \cdot \nabla v_h \approx - \sum_{i < j} \delta_E(\phi_h) \delta_E(v_h) \tilde{\varepsilon}_E(\mathbf{b}) (a_{i,j}^K).$$

De forma equivalente se escribe:

$$\int_K J(u_h) \cdot \nabla v_h \approx - \sum_{i < j} \tilde{\varepsilon}_E(\mathbf{b}) (a_{i,j}^K) \delta_E(\phi_h) \delta_E(v_h).$$

Y dado (3.18), se infiere lo siguiente:

$$\int_K J(u_h) \cdot \nabla v_h \approx \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h).$$

Recuerde que nuestra forma bilineal es:

$$a(u_h, v_h) = \int_K J(u_h) \cdot \nabla v_h + \int_K \gamma u_h v_h.$$

Así

$$\int_K J(u_h) \cdot \nabla v_h + \int_K \gamma u_h v_h \approx \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h) + \int_K \gamma u_h v_h. \quad (3.29)$$

Para el último término del lado derecho se utilizó la aproximación de Mass Lumping es decir,

$$\gamma_K(u_h, v_h) := \frac{|K|}{n+1} \sum_{i=1}^{n+1} \gamma(q_i) u_h(q_i) v_h(q_i) \approx \int_K \gamma u_h v_h. \quad (3.30)$$

Reemplazando (3.30) en (3.29) se obtiene,

$$\int_K J(u_h) \cdot \nabla v_h + \int_K \gamma u_h v_h \approx \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h) + \gamma_K(u_h, v_h).$$

Y sumando sobre todos los elementos de la triangulación, se tiene:

$$\sum_{K \in \mathcal{T}_h} \int_K J(u_h) \cdot \nabla v_h + \sum_{K \in \mathcal{T}_h} \int_K \gamma u_h v_h \approx \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h) + \gamma_K(u_h, v_h) \right\}.$$

Por tanto se obtiene:

$$\int_{\Omega} J(u_h) \cdot \nabla v_h + \int_{\Omega} \gamma u_h v_h \approx \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h) + \gamma_K(u_h, v_h) \right\}.$$

Finalmente se define la forma bilineal

$$a_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \tilde{\varepsilon}_E(\mathbf{b}) \omega_E^K \delta_E(e^\psi u_h) \delta_E(v_h) + \gamma_K(u_h, v_h) \right\}.$$

Observación 3.2. Note que $a_h(u_h, v_h)$ es una aproximación de $a(u_h, v_h)$.

Así cuando $\mathbf{b} = \nabla c$ con $c \in C^1(\Omega)$ el **problema discreto** definido por:

$$\begin{cases} \text{Hallar } u_h \in V_h \text{ tal que} \\ a_h(u_h, v_h) = F(v_h), \quad \forall v_h \in V_h. \end{cases} \quad (3.31)$$

Siguiendo con la idea del método, se analiza también el caso más general cuando \mathbf{b} no es el gradiente de alguna función continua, para ello se siguen las ideas de [5] y se tiene en cuenta los siguientes resultados.

DEFINICIÓN 3.2. Dado cualquier borde E , se introduce una función ψ_E definido localmente en E por la relación

$$\frac{\partial \psi_E}{\partial \tau_E} = \frac{1}{|\tau_E|} \varepsilon^{-1}(\mathbf{b} \cdot \tau_E). \quad (3.32)$$

donde $\partial/\partial \tau_E$ denota la derivada tangencial a lo largo del borde E .

El siguiente lema es de gran utilidad para detallar la derivación del método de [5].

LEMA 3.4. Sea $v \in H_0^1(\Omega) \cap C^0(\bar{\Omega})$ y $J(v) = \varepsilon \nabla v + \mathbf{b}v$. Entonces

$$\delta_E(e^{\psi_E} v) = \frac{1}{|\tau_E|} \int_E \varepsilon^{-1} e^{\psi_E} (J(v) \cdot \tau_E) ds. \quad (3.33)$$

Demostración. La demostración se sigue como en [5]. Sea

$$(J(v), \tau_E) = (\varepsilon \nabla v + \mathbf{b}v, \tau_E). \quad (3.34)$$

Multiplicando ε^{-1} en (3.34) se tiene,

$$\begin{aligned} (\varepsilon^{-1} J(v), \tau_E) &= (\nabla v + \varepsilon^{-1} \mathbf{b}v, \tau_E) \\ &= (\nabla v, \tau_E) + \varepsilon^{-1} (\mathbf{b}v, \tau_E). \end{aligned} \quad (3.35)$$

Dividiendo $|\tau_E|$ en (3.35) se reescribe de forma equivalente,

$$\frac{1}{|\tau_E|} (\nabla v, \tau_E) + \frac{\varepsilon^{-1}}{|\tau_E|} (\mathbf{b}v, \tau_E) = \frac{1}{|\tau_E|} (\varepsilon^{-1} J(v), \tau_E). \quad (3.36)$$

A su vez utilizando la igualdad (3.32) y de la definición de derivada direccional se tiene:

$$\frac{1}{|\tau_E|} (\nabla v, \tau_E) = \frac{\partial v}{\partial \tau_E}, \quad \frac{\varepsilon^{-1}}{|\tau_E|} (\mathbf{b}v, \tau_E) = \left(\frac{\partial \psi_E}{\partial \tau_E} \right) v. \quad (3.37)$$

Reemplazando (3.37) en (3.36) resulta:

$$\left(\frac{\partial v}{\partial \tau_E} + \left(\frac{\partial \psi_E}{\partial \tau_E} \right) v \right) = \frac{1}{|\tau_E|} \varepsilon^{-1} (J(v), \tau_E). \quad (3.38)$$

Multiplicando e^{ψ_E} en (3.38) se tiene,

$$e^{\psi_E} \left(\frac{\partial v}{\partial \tau_E} + e^{\psi_E} \left(\frac{\partial \psi_E}{\partial \tau_E} \right) v \right) = \frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} (J(v), \tau_E). \quad (3.39)$$

Por definición de derivada se tiene que:

$$\frac{\partial}{\partial \tau_E} (e^{\psi_E} v) = e^{\psi_E} \left(\frac{\partial v}{\partial \tau_E} + e^{\psi_E} \left(\frac{\partial \psi_E}{\partial \tau_E} \right) v \right).$$

Reemplazando la igualdad anterior en la ecuación (3.39) se obtiene lo siguiente,

$$\frac{\partial}{\partial \tau_E} (e^{\psi_E} v) = \frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} (J(v), \tau_E). \quad (3.40)$$

E integrando sobre el borde $E = q_i - q_j$, con q_i, q_j vértices del elemento K entonces,

$$\int_E \frac{\partial}{\partial \tau_E} (e^{\psi_E} v) = \frac{1}{|\tau_E|} \int_E \varepsilon^{-1} e^{\psi_E} (J(v), \tau_E). \quad (3.41)$$

Como E es la arista que conecta los vértices q_i y q_j , observe que,

$$\int_E \frac{\partial}{\partial \tau_E} (e^{\psi_E} v) = e^{\psi_E(q_i)} v(q_i) - e^{\psi_E(q_j)} v(q_j). \quad (3.42)$$

De la igualdad (3.15) se tiene:

$$\delta_E(e^{\psi_E} v) = e^{\psi_E(q_i)} v(q_i) - e^{\psi_E(q_j)} v(q_j). \quad (3.43)$$

Reemplazando (3.43) en (3.42) resulta,

$$\int_E \frac{\partial}{\partial \tau_E} (e^{\psi_E} v) = \delta_E(e^{\psi_E} v). \quad (3.44)$$

Finalmente reemplazando (3.44) en (3.41),

$$\delta_E(e^{\psi_E} v) = \frac{1}{|\tau_E|} \int_E \varepsilon^{-1} e^{\psi_E} (J(v) \cdot \tau_E) ds.$$

□

Observación 3.3. Asumiendo las siguientes aproximaciones

$$\left[\frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right] \approx \left[\frac{1}{|\tau_E|} \int_E \frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right], \quad (3.45)$$

$$\left[\frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} \right]^{-1} \approx \left[\frac{1}{|\tau_E|} \int_E \frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} \right]^{-1}. \quad (3.46)$$

De (3.44) y de la definición de promedio armónico $\varepsilon e^{-\psi_E}$ sobre la arista E recuerde que:

$$\int_E \frac{\partial(e^{\psi_E} v)}{\partial \tau_E} = \delta_E(e^{\psi_E} v), \quad \text{y} \quad \left[\frac{1}{|\tau_E|} \int_E \varepsilon^{-1} e^{\psi_E} \right]^{-1} = \tilde{\varepsilon}_E(\mathbf{b}).$$

Si multiplicamos los lados derechos de las aproximaciones (3.45) y (3.46) resulta,

$$\begin{aligned} \left[\frac{1}{|\tau_E|} \int_E \frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right] \left[\frac{1}{|\tau_E|} \int_E \frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} \right]^{-1} &= \left[\int_E \frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right] \left[\frac{1}{|\tau_E|} \int_E \varepsilon^{-1} e^{\psi_E} \right]^{-1} \\ &= \delta_E(e^{\psi_E} v) \tilde{\varepsilon}_E(\mathbf{b}). \end{aligned}$$

Observe que si multiplicamos los lados izquierdos de (3.45) y (3.46) resulta,

$$\left[\frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right] \left[\frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} \right]^{-1} \approx \delta_E(e^{\psi_E} v) \tilde{\varepsilon}_E(\mathbf{b}). \quad (3.47)$$

Además de (3.40) se tiene,

$$J(v) \cdot \tau_E = \left[\frac{\partial(e^{\psi_E} v)}{\partial \tau_E} \right] \left[\frac{1}{|\tau_E|} \varepsilon^{-1} e^{\psi_E} \right]^{-1}. \quad (3.48)$$

Por tanto reemplazando (3.47) en (3.48) se infiere lo siguiente:

$$J(v) \cdot \tau_E \approx \delta_E(e^{\psi_E} v) \tilde{\varepsilon}_E(\mathbf{b}).$$

Regresando a la derivación del esquema propuesto en [5], se tiene en cuenta la igualdad (3.19), entonces para cualquier $v_h \in V_h$, se obtiene,

$$\int_K J(u_h) \cdot \nabla v_h = \sum_{E \subset K} \omega_E^K (J(u_h) \cdot \tau_E) \delta_E(v_h) \approx \sum_{E \subset K} \omega_E^K \delta_E(e^{\psi_E} u_h) \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(v_h).$$

Y procediendo de forma similar que en el caso anterior se infiere lo siguiente,

$$\int_K J(u_h) \cdot \nabla v_h \approx \sum_{E \subset K} \omega_E^K \delta_E(e^{\psi_E} u_h) \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(v_h). \quad (3.49)$$

Utilizando la aproximación de Mass Lamping.

$$\int_K J(u_h) \cdot \nabla v_h + \int_K \gamma u_h v_h \approx \sum_{E \subset K} \omega_E^K \delta_E(e^{\psi_E} u_h) \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(v_h) + \gamma_K(u_h, v_h).$$

Y sumando sobre todos los elementos de la triangulación resulta

$$\int_{\Omega} J(u_h) \cdot \nabla v_h + \int_{\Omega} \gamma u_h v_h \approx \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \omega_E^K \delta_E(e^{\psi_E} u_h) \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(v_h) + \gamma_K(u_h, v_h) \right\}.$$

Finalmente se define la forma bilineal

$$a_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \omega_E^K \delta_E(e^{\psi_E} u_h) \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(v_h) + \gamma_K(u_h, v_h) \right\}. \quad (3.50)$$

De esta forma, cuando \mathbf{b} no es el gradiente de alguna función continua el **problema discreto** es:

$$\begin{cases} \text{Hallar } u_h \in V_h \text{ tal que} \\ a_h(u_h, v_h) = F(v_h), \quad \forall v_h \in V_h. \end{cases} \quad (3.51)$$

De los resultados dados anteriores, $\forall u \in V \cap C^0(\bar{\Omega})$, se deduce la siguiente afirmación (ver [5]):

$$a_h(u_I, v_h) := \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(u) \cdot \tau_E \right] \delta_E(v_h) + \gamma_K(u_I, v_h) \right\}, \quad (3.52)$$

donde $u_I \in V_h$ es el interpolante de Lagrange de u .

El siguiente resultado de [5] nos dice que el esquema discreto (3.51) es un esquema monótono y la demostración del resultado se sigue como en [5]

LEMA 3.5. *La correspondiente matriz de rigidez de la forma bilineal (3.50) es una M -matriz para cualquier función continua $\varepsilon > 0$ y \mathbf{b} si y sólo si la matriz de rigidez de la ecuación de Poisson es una M -matriz. Es decir si y sólo si satisface la condición (3.20).*

Demostración. Dado $j \in \{1, \dots, N_h\}$ se considera el correspondiente nodo x_j . Obviamente si x_i es vecino de x_j entonces,

$$A_{ij} = \sum_{x_j \in E} \omega_E \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(e^{\psi_E} \varphi_j) \delta_E(\varphi_i) = -\omega_E \tilde{\varepsilon}_E(\mathbf{b}) (e^{\psi_{j,E}}) \leq 0.$$

Aquí $x_j \in E$, significa el borde E que contiene x_j como punto final, y $\psi_{j,E} = \psi_E(x_j)$. Ahora si x_j no tiene nodo vecino en la frontera, entonces la suma de la j -ésima columna de A es cero:

$$\begin{aligned} \sum_i A_{ij} &= \sum_{x_j \in E} \omega_E \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(e^{\psi_E} \varphi_j) \delta_E \sum_i \varphi_i \\ &= \sum_{x_j \in E} \omega_E \tilde{\varepsilon}_E(\mathbf{b}) \delta_E(e^{\psi_E} \varphi_j) \delta_E 1 = 0. \end{aligned}$$

lo que significa que $A_{jj} = \sum_{i \neq j} |A_{ij}|$. Y si x_j tiene un nodo vecino en la frontera, es fácil ver que,

$$\sum_i A_{ij} > 0 \quad \text{o} \quad A_{jj} > \sum_{i \neq j} A_{ij}.$$

Lo que concluye la demostración. □

En lo que sigue demostraremos que el esquema discreto EAFEM (3.51) tiene una única solución. Para ello suponemos que $\forall K \in \mathcal{T}_h$, la solución del problema (2.1) satisface $J(u) = (\varepsilon \nabla u + \mathbf{b}u) \in [W^{1,p}(K)]^d$, $\gamma \in C(\bar{K})$, $\gamma u \in W^{1,r}(K)$ con $r > d$,

$$p = 2 \text{ para } d = 2 \text{ y } p > d - 1 \text{ para } d > 2. \quad (3.53)$$

En adelante se estudiará el caso cuando $d = 2$ y $p = 2$.

El primer paso para la demostración de la existencia y unicidad de la solución del problema discreto EAFEM es dar una estimación como en [5] de la diferencia sobre las formas bilineal continua $a(\cdot, \cdot)$ y la forma bilineal discreta $a_h(\cdot, \cdot)$ en el siguiente lema, la demostración del resultado se sigue como en [5].

LEMA 3.6. Sea $w \in C(\bar{\Omega})$, si $\forall K \in \mathcal{T}_h$ se tiene $J(w) \in [H^1(K)]^2$, $\gamma w \in W^{1,r}(K)$, entonces la siguiente desigualdad sostiene que para cada $v_h \in V_h$:

$$|a(w, v_h) - a_h(w_I, v_h)| \leq ch \left\{ \sum_{K \in \mathcal{T}_h} |J(w)|_{1,K}^2 + \sum_{K \in \mathcal{T}_h} |\gamma w|_{1,r,K}^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega},$$

donde c es una constante independiente de la medida de h .

Demostración. De la definición de las formas bilineales $a(\cdot, \cdot)$ y $a_h(\cdot, \cdot)$, se tiene:

$$\begin{aligned} a(w, v_h) - a_h(w_I, v_h) &= \int_{\Omega} (\varepsilon \nabla w + \mathbf{b}w) \cdot \nabla v_h + \int_{\Omega} \gamma w v_h \\ &\quad - \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(w) \cdot \tau_E \right] \delta_E(v_h) + \gamma_K(w_I, v_h) \right\}. \end{aligned}$$

Teniendo en cuenta que Ω se puede escribir como una suma de elementos de la triangulación \mathcal{T}_h resulta,

$$\begin{aligned} a(w, v_h) - a_h(w_I, v_h) &= \sum_{K \in \mathcal{T}_h} \left\{ \int_K (\varepsilon \nabla w + \mathbf{b}w) \cdot \nabla v_h + \int_K \gamma w v_h \right\} \\ &\quad - \sum_{K \in \mathcal{T}_h} \left\{ \sum_{E \subset K} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(w) \cdot \tau_E \right] \delta_E(v_h) + \gamma_K(w_I, v_h) \right\}. \end{aligned}$$

De forma equivalente podemos escribir

$$\begin{aligned} a(w, v_h) - a_h(w_I, v_h) &= \sum_{K \in \mathcal{T}_h} \left\{ \int_K J(w) \cdot \nabla v_h - \sum_{E \subset K} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(w) \cdot \tau_E \right] \delta_E(v_h) \right\} \\ &\quad + \sum_{K \in \mathcal{T}_h} \left\{ \int_K (\gamma w v_h - \gamma_K(w_I, v_h)) \right\}. \end{aligned}$$

Haciendo

$$\begin{aligned} \mathcal{E}_K(J(w), v_h) &= \int_K J(w) \cdot \nabla v_h - \sum_{E \subset K} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(w) \cdot \tau_E \right] \delta_E(v_h), \\ \mathcal{Q}_K(\gamma w, v_h) &= \int_K (\gamma w v_h - \gamma_K(w_I, v_h)), \end{aligned}$$

se obtiene,

$$a(w, v_h) - a_h(w_I, v_h) = \sum_{K \in \mathcal{T}_h} \mathcal{E}_K(J(w), v_h) + \sum_{K \in \mathcal{T}_h} \mathcal{Q}_K(\gamma w, v_h).$$

Se empieza la demostración acotando $\mathcal{E}_K(J(w), v_h)$ en el elemento de referencia \hat{K} ,

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| = \left| \int_{\hat{K}} J(w) \cdot \nabla v_h - \sum_{E \subset \hat{K}} \omega_E^K \left[\frac{\tilde{\varepsilon}_E(\mathbf{b})}{|\tau_E|} \int_E \frac{e^{\psi_E}}{\varepsilon} J(w) \cdot \tau_E \right] \delta_E(v_h) \right| \quad (3.54)$$

$$\leq \left| \int_{\hat{K}} \widehat{J(w)} \cdot \nabla \hat{v}_h \right| + \left| \sum_{\hat{E} \subset \hat{K}} \omega_{\hat{E}}^{\hat{K}} \left[\frac{\tilde{\varepsilon}_{\hat{E}}(\mathbf{b})}{|\tau_{\hat{E}}|} \int_{\hat{E}} \frac{e^{\psi_{\hat{E}}}}{\varepsilon} \widehat{J(w)} \cdot \tau_{\hat{E}} \right] \delta_{\hat{E}}(\hat{v}_h) \right|, \quad (3.55)$$

utilizando desigualdad de Cauchy Schwarz, se deduce:

$$\begin{aligned} |\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| &\leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\nabla \hat{v}_h\|_{0,\hat{K}} + \sum_{\hat{E} \subset \hat{K}} \frac{|\omega_{\hat{E}}^{\hat{K}} \tilde{\varepsilon}_{\hat{E}}(\mathbf{b})|}{|\tau_{\hat{E}}|} \left| \int_{\hat{E}} \frac{e^{\psi_{\hat{E}}}}{\varepsilon} \widehat{J(w)} \cdot \tau_{\hat{E}} \right| |\delta_{\hat{E}}(\hat{v}_h)| \\ &\leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\nabla \hat{v}_h\|_{0,\hat{K}} + \sum_{\hat{E} \subset \hat{K}} \frac{|\omega_{\hat{E}}^{\hat{K}} \tilde{\varepsilon}_{\hat{E}}(\mathbf{b})|}{|\tau_{\hat{E}}|} \left\| \frac{e^{\psi_{\hat{E}}} \tau_{\hat{E}}}{\varepsilon} \right\|_{0,\infty,\hat{E}} \|\widehat{J(w)}\|_{0,1,\hat{E}} |\delta_{\hat{E}}(\hat{v}_h)|. \end{aligned}$$

Observando que

$$\delta_{\hat{E}}(\hat{v}_h) = \hat{v}_h(q_i) - \hat{v}_h(q_j) = \int_{\hat{E}} \frac{\partial \hat{v}_h}{\partial q_i} = \left(\frac{\partial \hat{v}_h}{\partial q_i}, 1 \right)_{0,\hat{E}}.$$

Utilizando la Desigualdad de Cauchy Schwarz resulta,

$$|\delta_{\hat{E}}(\hat{v}_h)| \leq \|1\|_{0,\hat{E}} \left\| \frac{\partial \hat{v}_h}{\partial x_i} \right\| \leq (\hat{E})^{\frac{1}{2}} \|\nabla \hat{v}_h\|_{0,\hat{E}},$$

así se obtiene,

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| \leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\nabla \hat{v}_h\|_{0,\hat{K}} + \sum_{\hat{E} \subset \hat{K}} \frac{|\omega_{\hat{E}}^{\hat{K}} \tilde{\varepsilon}_{\hat{E}}(\mathbf{b})|}{|\tau_{\hat{E}}|} \left\| \frac{e^{\psi_{\hat{E}}} \tau_{\hat{E}}}{\varepsilon} \right\|_{0,\infty,\hat{E}} \|\widehat{J(w)}\|_{0,1,\hat{E}} (\hat{E})^{\frac{1}{2}} \|\nabla \hat{v}_h\|_{0,\hat{E}}.$$

Y haciendo,

$$c = \frac{|\omega_{\hat{E}}^{\hat{K}} \tilde{\varepsilon}_{\hat{E}}(\mathbf{b})|}{|\tau_{\hat{E}}|} \left\| \frac{e^{\psi_{\hat{E}}} \tau_{\hat{E}}}{\varepsilon} \right\|_{0,\infty,\hat{E}} (\hat{E})^{\frac{1}{2}},$$

entonces se infiere,

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| \leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\nabla \hat{v}_h\|_{0,\hat{K}} + c \sum_{\hat{E} \subset \hat{K}} \|\widehat{J(w)}\|_{0,1,\hat{E}} \|\nabla \hat{v}_h\|_{0,\hat{E}}.$$

Teniendo en cuenta que $\sum_{\hat{E} \subset \hat{K}} \hat{E} = \partial \hat{K}$ resulta,

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| \leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\hat{v}_h\|_{1,\hat{K}} + c_0 \|\widehat{J(w)}\|_{0,1,\partial \hat{K}} \|\nabla \hat{v}_h\|_{0,\partial \hat{K}},$$

del Teorema de trazas, resulta $\|\widehat{J(w)}\|_{0,1,\partial\hat{K}} \leq \|\widehat{J(w)}\|_{1,\hat{K}}$, entonces

$$\begin{aligned} |\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| &\leq \|\widehat{J(w)}\|_{0,\hat{K}} \|\hat{v}_h\|_{1,\hat{K}} + c_0 \|\widehat{J(w)}\|_{1,\hat{K}} \|\nabla \hat{v}_h\|_{0,\partial\hat{K}} \\ &\leq \max\{1, c_0\} (\|\widehat{J(w)}\|_{0,\hat{K}} \|\hat{v}_h\|_{1,\hat{K}} + \|\widehat{J(w)}\|_{1,\hat{K}} \|\nabla \hat{v}_h\|_{0,\partial\hat{K}}) \\ &\leq \max\{1, c_0\} (\|\widehat{J(w)}\|_{1,\hat{K}} \|\hat{v}_h\|_{1,\hat{K}} + \|\widehat{J(w)}\|_{1,\hat{K}} \|\nabla \hat{v}_h\|_{0,\partial\hat{K}}). \end{aligned}$$

haciendo $c_1 = \max\{1, c_0\}$ y del teorema de Trazas se obtiene,

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| \leq c_1 \|\widehat{J(w)}\|_{0,\hat{K}} \|\hat{v}_h\|_{1,\hat{K}}.$$

Luego de la derivación del esquema EAFEM, es claro ver que $\mathcal{E}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h) = 0$, si $J(w) \equiv$ constante en \hat{K} y teniendo en cuenta el Lema de Bramble Hilbert en \hat{K} , es fácil ver que:

$$|\hat{\mathcal{E}}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h)| \leq Ch |\widehat{J(w)}|_{1,\hat{K}} |\hat{v}_h|_{1,\hat{K}}.$$

Volviendo del elemento de referencia \hat{K} a cualquier elemento K de la triangulación \mathcal{T}_h , notemos que:

$$|\mathcal{E}_K(J(w), v_h)| \leq Ch |J(w)|_{1,K} |v_h|_{1,K}.$$

La estimación para \mathcal{Q}_K se sigue de \mathcal{E}_K , (note que $\mathcal{Q}_{\hat{K}}(\widehat{J(w)}, \hat{v}_h) = 0$, si $J(w) \equiv$ constante en \hat{K}), obteniéndose

$$|\mathcal{Q}_K(\gamma w, v_h)| \leq Ch |\gamma w|_{1,r,K} \|v_h\|_{1,K}.$$

Finalmente de los resultados anteriores se sigue que,

$$\begin{aligned} |a(w, v_h) - a_h(w_I, v_h)| &\leq \sum_{K \in \mathcal{T}_h} Ch |J(w)|_{1,K} |v_h|_{1,K} + \sum_{K \in \mathcal{T}_h} Ch |\gamma w|_{1,r,K} \|v_h\|_{1,K} \\ &\leq Ch \sum_{K \in \mathcal{T}_h} \left\{ |J(w)|_{1,K} |v_h|_{1,K} + |\gamma w|_{1,r,K} \|v_h\|_{1,K} \right\} \\ &\leq Ch \sum_{K \in \mathcal{T}_h} (|J(w)|_{1,K}^2 + |\gamma w|_{1,r,K}^2)^{\frac{1}{2}} (|v_h|_{1,K}^2 + \|v_h\|_{1,K}^2)^{\frac{1}{2}} \\ &\leq Ch \left\{ \sum_{K \in \mathcal{T}_h} |J(w)|_{1,K}^2 + |\gamma w|_{1,r,K}^2 \right\}^{\frac{1}{2}} \left\{ \sum_{K \in \mathcal{T}_h} |v_h|_{1,K}^2 + \|v_h\|_{1,K}^2 \right\}^{\frac{1}{2}} \\ &\leq Ch \left\{ \sum_{K \in \mathcal{T}_h} |J(w)|_{1,K}^2 + |\gamma w|_{1,r,K}^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega}. \end{aligned}$$

□

Procediendo como en [5], el lema siguiente nos garantiza que el esquema discreto EAFEM (3.51) tiene una única solución. Además la demostración se sigue como en [5] pero generalizando la demostración para el caso cuando $\gamma \neq 0$.

LEMA 3.7. *Dados $\varepsilon \in W^{1,\infty}(K)$, $\gamma \in W^{1,\infty}(K)$ y $\mathbf{b} \in [W^{1,\infty}(K)]^2$, $\forall K \in \mathcal{T}_h$. Entonces para h suficientemente pequeño se tiene:*

$$\sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|v_h\|_{1,\Omega}} \geq c_0 \|w_h\|_{1,\Omega}, \quad \forall w_h \in V_h.$$

Demostración. Se sigue del Teorema 2.1 y del Teorema de Lax–Milgram Generalizado, que para h suficientemente pequeño,

$$\sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_{1,\Omega}} \geq 2c_0 \|w_h\|_{1,\Omega}, \quad \forall w_h \in V_h, \quad (3.56)$$

además para cualquier $v_h, w_h \in V_h$, podemos escribir,

$$a_h(w_h, v_h) = a(w_h, v_h) - [a(w_h, v_h) - a_h(w_h, v_h)]. \quad (3.57)$$

Siguiendo el Lema 3.6 se tiene,

$$|a(w_h, v_h) - a_h(w_h, v_h)| \leq Ch \left\{ \sum_{K \in \mathcal{T}_h} |J(w_h)|_{1,K}^2 + \sum_{K \in \mathcal{T}_h} |\gamma w_h|_{1,r,K}^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega}. \quad (3.58)$$

Observe que,

$$\begin{aligned} |J(w_h)|_{1,K}^2 &= \int_K |\nabla(\varepsilon \nabla w_h + \mathbf{b} w_h)|^2 = \int_K |\nabla(\varepsilon \nabla w_h + \mathbf{b} w_h) : \nabla(\varepsilon \nabla w_h + \mathbf{b} w_h)| \\ &= \int_K |\nabla(\varepsilon \nabla w_h) : \nabla(\varepsilon \nabla w_h)| + \int_K |\nabla(\varepsilon \nabla w_h) : \nabla(\mathbf{b} w_h)| \\ &\quad + \int_K |\nabla(\mathbf{b} w_h) : (\varepsilon \nabla w_h)| + \int_K |\nabla(\mathbf{b} w_h) : \nabla(\mathbf{b} w_h)| \\ &= \int_K |\nabla(\varepsilon \nabla w_h) : \nabla(\varepsilon \nabla w_h)| + |2\nabla(\varepsilon \nabla w_h) : \nabla(\mathbf{b} w_h)| + |\nabla(\mathbf{b} w_h) : \nabla(\mathbf{b} w_h)|. \end{aligned}$$

Además se sabe que $|w_h|_{2,K} = 0$, entonces para cualquier $w_h \in V$ y $K \in \mathcal{T}_h$, se tiene:

$$|J(w_h)|_{1,K} \leq C(\|\varepsilon\|_{1,\infty,K} + \|\mathbf{b}\|_{1,\infty,K}) \|w_h\|_{1,K}. \quad (3.59)$$

Reemplazando (3.59) en (3.58), es equivalente escribir,

$$\begin{aligned}
& |a(w_h, v_h) - a_h(w_h, v_h)| \\
& \leq Ch \left\{ \sum_{K \in \mathcal{T}_h} \left| C(\|\varepsilon\|_{1,\infty,K} + \|\mathbf{b}\|_{1,\infty,K}) \|w_h\|_{1,K} \right|^2 + \left| C\|\gamma\|_{1,\infty,K} \|w_h\|_{1,K} \right|^2 \right\}^{\frac{1}{2}} \|v_h\|_{1,\Omega} \\
& \leq Ch \max_{K \in \mathcal{T}_h} (\|\varepsilon\|_{1,\infty,K} + \|\mathbf{b}\|_{1,\infty,K} + \|\gamma\|_{1,\infty,K}) \|w_h\|_{1,\Omega} \|v_h\|_{1,\Omega}.
\end{aligned}$$

Eligiendo $h < h_0$,

$$|a(w_h, v_h) - a_h(w_h, v_h)| \leq Ch_0 \max_{K \in \mathcal{T}_h} (\|\varepsilon\|_{1,\infty,K} + \|\mathbf{b}\|_{1,\infty,K} + \|\gamma\|_{1,\infty,K}) \|w_h\|_{1,\Omega} \|v_h\|_{1,\Omega}.$$

En particular se elige $h_0 = c_0 \left(C \max_{K \in \mathcal{T}_h} (\|\varepsilon\|_{1,\infty,K} + \|\mathbf{b}\|_{1,\infty,K} + \|\gamma\|_{1,\infty,K}) \right)^{-1}$ resulta,

$$|a(w_h, v_h) - a_h(w_h, v_h)| \leq c_0 \|w_h\|_{1,\Omega} \|v_h\|_{1,\Omega}. \quad (3.60)$$

Multiplicando por -1 en ambos lados de la ecuación se obtiene,

$$-|a(w_h, v_h) - a_h(w_h, v_h)| \geq -c_0 \|w_h\|_{1,\Omega} \|v_h\|_{1,\Omega}. \quad (3.61)$$

Por último regresando a (3.57) y tomando supremo en V_h resulta,

$$\begin{aligned}
\sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|v_h\|_{1,\Omega}} & \geq \sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_{1,\Omega}} - \sup_{v_h \in V_h} \frac{|a(w_h, v_h) - a_h(w_h, v_h)|}{\|v_h\|_{1,\Omega}} \\
& \geq \sup_{v_h \in V_h} 2c_0 \|w_h\|_{1,\Omega} - \sup_{v_h \in V_h} c_0 \|w_h\|_{1,\Omega} \\
& = c_0 \|w_h\|_{1,\Omega}.
\end{aligned}$$

Por lo tanto se concluye la demostración del Lema. □

Del Lema anterior y del Teorema de Lax–Milgram Generalizado, se tiene una estimación apriori para la solución del esquema discreto EAFEM.

LEMA 3.8. *Existe un único $u_h \in V_h$ solución del problema (3.51) tal que:*

$$c_0 \|u_h\|_{1,\Omega} \leq \|f\|_{0,\Omega} + \|g\|_{0,\Gamma_N}.$$

Demostración. Se sabe que,

$$a_h(u_h, v_h) = F(v_h).$$

Entonces se obtiene,

$$c_0 \|u_h\|_{1,\Omega} \leq \sup_{v_h \in V_h} \frac{a_h(u_h, v_h)}{\|v_h\|_{1,\Omega}} = \sup_{v_h \in V_h} \frac{F(v_h)}{\|v_h\|_{1,\Omega}} \leq \|f\|_{0,\Omega} + \|g\|_{0,\Gamma_N}.$$

□

Como consecuencia de los resultados anteriores se tiene una estimación apriori para el error en el siguiente teorema, y la demostración se sigue como en [5].

TEOREMA 3.1. *Sea u la solución del problema continuo. Asumimos que para todo $K \in \mathcal{T}_h$, se tiene $\varepsilon \in W^{1,\infty}(K)$, $\beta \in [W^{1,\infty}(K)]^2$, $J(u) = (\varepsilon \nabla u + \beta u) \in [H^1(K)]^2$, $\gamma \in C(\bar{K})$, $\gamma u \in W^{1,r}(K)$. Entonces se tiene la siguiente estimación para h suficientemente pequeño,*

$$\|u_I - u_h\|_{1,\Omega} \leq ch \left\{ \sum_{K \in \mathcal{T}_h} |J(u)|_{1,K}^2 + \sum_{K \in \mathcal{T}_h} |\gamma u|_{1,r,K}^2 \right\}^{\frac{1}{2}}. \quad (3.62)$$

Demostración. Usando el Lema 3.7 se sigue que,

$$c_0 \|u_I - u_h\|_{1,\Omega} \leq \sup_{v_h \in V_h} \frac{a_h(u_I - u_h, v_h)}{\|v_h\|_{1,\Omega}},$$

la desigualdad anterior es equivalente a,

$$\begin{aligned} c_0 \|u_I - u_h\|_{1,\Omega} &\leq \sup_{v_h \in V_h} \frac{|a_h(u_I, v_h) - a_h(u_h, v_h)|}{\|v_h\|_{1,\Omega}} \\ &= \sup_{v_h \in V_h} \frac{|a_h(u_I, v_h) - a_h(u_h, v_h) + a(u, v_h) - a(u, v_h)|}{\|v_h\|_{1,\Omega}} \\ &= \sup_{v_h \in V_h} \frac{|a(u, v_h) - a_h(u_h, v_h) - [a(u, v_h) - a_h(u_I, v_h)]|}{\|v_h\|_{1,\Omega}} \\ &\leq \sup_{v_h \in V_h} \frac{|a(u, v_h) - a_h(u_I, v_h)|}{\|v_h\|_{1,\Omega}} + \sup_{v_h \in V_h} \frac{|a(u, v_h) - a_h(u_h, v_h)|}{\|v_h\|_{1,\Omega}}. \end{aligned}$$

observando que $a(u, v_h) = F(v_h)$ y $a_h(u_h, v_h) = F(v_h)$, entonces

$$a(u, v_h) - a_h(u_h, v_h) = 0.$$

De donde se obtiene lo siguiente,

$$c_0 \|u_I - u_h\|_{1,\Omega} \leq \sup_{v_h \in V_h} \frac{|a(u, v_h) - a_h(u_I, v_h)|}{\|v_h\|_{1,\Omega}}.$$

Por último utilizando el Lema 3.6 se concluye la demostración, es decir,

$$\|u_I - u_h\|_{1,\Omega} \leq ch \left\{ \sum_{K \in \mathcal{T}_h} |J(u)|_{1,K}^2 + \sum_{K \in \mathcal{T}_h} |\gamma u|_{1,r,K}^2 \right\}^{\frac{1}{2}}.$$

□

3.3. Ejemplos numéricos

En esta sección se desarrolla un ejemplo numérico que ilustra el buen funcionamiento y difusividad del esquema de elementos finitos EAFEM (3.51), y lo ineficiente que es el método de Galerkin para resolver las ecuaciones de convección–difusión–reacción con convección dominada. En lo que sigue \mathbf{b} representará el término convectivo, ε el término difusivo, γ el término de reacción y N_h representará el número total de grados de libertad o nodos de la malla que se define a partir de el espacio de dimensión finita V_h , con h que representa la medida de la malla uniforme.

3.3.1. Ejemplo 1

En el siguiente ejemplo numérico, la solución de la ecuación (2.1) de convección–difusión–reacción del tipo estacionario tiene capa límite y sirve como motivación para hacer una comparación entre el esquema de Galerkin visto en la Sección 3.1 y el esquema EAFEM visto en la Sección 3.2. Para el cálculo de las soluciones aproximadas utilizaremos un código Matlab basado en la implementación numérica de [9] y [10], mostrando también las gráficas de las soluciones para ambos esquemas

En este ejemplo se consideran los siguientes datos:

- $\Omega = [0, 1] \times [0, 1] \subset \mathbb{R}^2$.
- $\partial\Omega = \Gamma = \{(x, 0) \in \mathbb{R}^2 : 0 \leq x \leq 1\} \cup \{(0, y) \in \mathbb{R}^2 : 0 \leq y \leq 1\}$.
- \mathbf{n} es el vector normal unitario exterior a la frontera.
- $u(x) = 1 - \frac{\sinh(e^{-1/2}x)}{\sinh(e^{-1/2})}$.
- Parámetros: $\varepsilon = 10^{-4}$, $\mathbf{b} = [0 \ 0]$, $\gamma = 1$.
- Lado derecho: $f = 1$.
- Condiciones de contorno: Dirichlet.

Para resolver la ecuación diferencial parcial elíptica de convección-difusión-reacción del tipo estacionario, en forma de divergencia y con condiciones de contorno Dirichlet homogéneo y Neumann no homogéneo, dada por,

$$\left\{ \begin{array}{ll} \mathbf{L}u := -\nabla \cdot (\varepsilon(x)\nabla u(x) + \mathbf{b}(x)u(x)) + \gamma(x)u(x) &= f(x), & x \in \Omega, \\ u(x) &= 0, & x \in \Gamma_D, \\ (\varepsilon(x)\nabla u(x) + \mathbf{b}(x)u(x)) \cdot \mathbf{n} &= g, & x \in \Gamma_N, \end{array} \right.$$

de manera computacional, se analiza el caso más simple, es decir se empieza dividiendo el dominio $\Omega = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ en la malla uniforme \mathcal{T}_h , donde la medida del diámetro de cada elemento es $h = 0,5$, teniendo como resultado 9 vértices los que denotaremos como x_i , con $i = \{1, \dots, 9\}$ y la matriz formada por las coordenadas de los vértices la llamaremos `vert`. De la partición del dominio también se obtienen 8 elementos o triángulos, y la matriz que contiene los vértices de los elementos la denominaremos `ele`. Para tener una idea más clara de la partición del dominio ver la Figura 3.3

$$\text{vert} = \begin{pmatrix} 0 & 0 \\ 0 & 0,5 \\ 0 & 1 \\ 0,5 & 0 \\ 0,5 & 0,5 \\ 0,5 & 1 \\ 1 & 0 \\ 1 & 0,5 \\ 1 & 1 \end{pmatrix}, \quad \text{ele} = \begin{pmatrix} 4 & 5 & 1 \\ 5 & 6 & 2 \\ 7 & 8 & 4 \\ 8 & 9 & 5 \\ 2 & 1 & 5 \\ 3 & 2 & 6 \\ 5 & 4 & 8 \\ 6 & 5 & 9 \end{pmatrix}$$

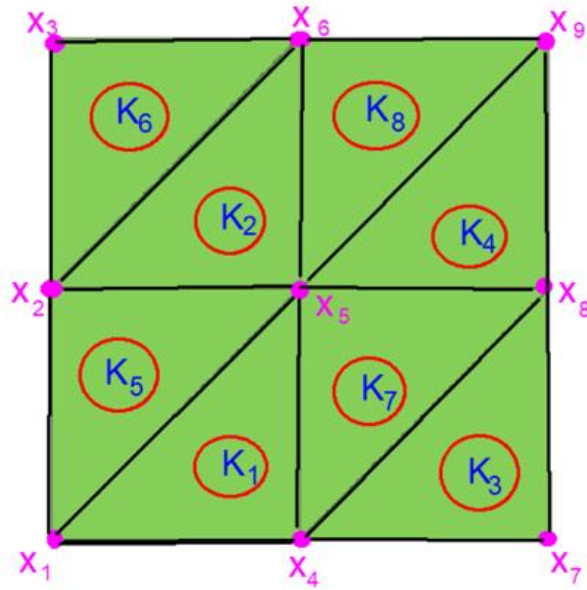


Figura 3.3: Malla con 9 grados de libertad y 8 elementos

Luego, se calcula la matriz de rigidez teniendo en cuenta la forma bilineal $a(\cdot, \cdot)$ visto en la Sección 2.2 y el espacio de dimensión finita formado por las funciones lineales y continuas en cada elemento, dada por $V_h =: \{v_h \in C^0(\bar{\Omega}) \cap H_D^1(\Omega) : v_h|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h\} \subset H_D^1(\Omega)$. Así, se tiene el siguiente problema discreto:

Hallar $u_h \in V_h$ **tal que**:

$$a(u_h, \varphi_i) = F(\varphi_i), \quad \forall i = \{1, \dots, 9\}. \quad (3.63)$$

Donde $\{\varphi_1, \dots, \varphi_9\}$ es una base de V_h . Como V_h es un espacio de dimensión finita, existen $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_9 \in \mathbb{R}$, tales que la solución u_h se puede escribir como combinación lineal de las incógnitas por las funciones bases, es decir:

$$u_h = \sum_{j=1}^9 \alpha_j \varphi_j = \alpha_1 \varphi_1 + \alpha_2 \varphi_2 + \alpha_3 \varphi_3 + \dots + \alpha_9 \varphi_9,$$

por lo tanto, el problema discreto (3.63) se puede reescribir como: Hallar $\alpha_1, \dots, \alpha_9 \in \mathbb{R}$ tal que:

$$\sum_{j=1}^n \alpha_j a(\varphi_j, \varphi_i) = F(\varphi_i), \quad \forall i = \{1, \dots, 9\}.$$

reescrito en su forma matricial resulta,

$$\underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_9 \end{bmatrix}}_{\vec{\alpha}} \underbrace{\begin{bmatrix} a(\varphi_1, \varphi_1) & a(\varphi_2, \varphi_1) & a(\varphi_3, \varphi_1) & \cdots & a(\varphi_9, \varphi_1) \\ a(\varphi_1, \varphi_2) & a(\varphi_2, \varphi_2) & a(\varphi_3, \varphi_2) & \cdots & a(\varphi_9, \varphi_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a(\varphi_1, \varphi_9) & a(\varphi_2, \varphi_9) & a(\varphi_3, \varphi_9) & \cdots & a(\varphi_9, \varphi_9) \end{bmatrix}}_{A1} = \underbrace{\begin{bmatrix} F(\varphi_1) \\ F(\varphi_2) \\ \vdots \\ F(\varphi_9) \end{bmatrix}}_{\mathbf{b}}.$$

De esta forma, la solución puede ser calculada como:

$$\underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_9 \end{bmatrix}}_{\vec{\alpha}} = \underbrace{\begin{bmatrix} F(\varphi_1) \\ F(\varphi_2) \\ \vdots \\ F(\varphi_9) \end{bmatrix}}_{\mathbf{b}} \cdot \underbrace{\begin{bmatrix} a(\varphi_1, \varphi_1) & a(\varphi_2, \varphi_1) & a(\varphi_3, \varphi_1) & \cdots & a(\varphi_9, \varphi_1) \\ a(\varphi_1, \varphi_2) & a(\varphi_2, \varphi_2) & a(\varphi_3, \varphi_2) & \cdots & a(\varphi_9, \varphi_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a(\varphi_1, \varphi_9) & a(\varphi_2, \varphi_9) & a(\varphi_3, \varphi_9) & \cdots & a(\varphi_9, \varphi_9) \end{bmatrix}^{-1}}_{A1},$$

así resulta el siguiente problema discreto,

$$\mathbf{Hallar} \vec{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_9 \end{pmatrix} \in \mathbb{R}^9 \mathbf{talque} \vec{\alpha} = \mathbf{b} A1^{-1}$$

Para encontrar la matriz A1, usamos la forma bilineal $a(\cdot, \cdot)$ y el mallado es decir,

$$\begin{aligned}
 \int_{\Omega} a(\varphi_i, \varphi_j) &= \int_{\Omega} (\varepsilon \nabla \varphi_i + \mathbf{b} \varphi_i) \cdot \nabla \varphi_j \int_{\Omega} \gamma \varphi_i \varphi_j \\
 &= \int_{\Omega} \varepsilon \nabla \varphi_i \cdot \nabla \varphi_j + \int_{\Omega} \mathbf{b} \varphi_i \nabla \varphi_j + \int_{\Omega} \gamma \varphi_i \varphi_j \\
 &= \int_{\text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)} \varepsilon \nabla \varphi_i \cdot \nabla \varphi_j + \int_{\text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)} \mathbf{b} \varphi_i \nabla \varphi_j + \int_{\text{sop}(\varphi_j) \cap \text{sop}(\varphi_j)} \gamma \varphi_i \varphi_j \\
 &= \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)}} \left(\underbrace{\int_K \varepsilon \nabla \varphi_i \cdot \nabla \varphi_j}_{M_{ij}} + \underbrace{\int_K \mathbf{b} \varphi_i \nabla \varphi_j}_{E_{ij}} + \underbrace{\int_K \gamma \varphi_i \varphi_j}_{M2_{ij}} \right).
 \end{aligned}$$

Dado que, ε se considera una función constante, resulta:

$$\int_{\Omega} a(\varphi_i, \varphi_j) = \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)}} \left(\varepsilon \underbrace{\int_K \nabla \varphi_i \cdot \nabla \varphi_j}_{M_{ij}} + \underbrace{\int_K \mathbf{b} \varphi_i \nabla \varphi_j}_{E_{ij}} + \underbrace{\int_K \gamma \varphi_i \varphi_j}_{M2_{ij}} \right).$$

Por lo tanto

$$\begin{aligned}
 M_{ij} &= \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)}} \int_K \nabla \varphi_i \cdot \nabla \varphi_j \\
 E_{ij} &= \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)}} \int_K \mathbf{b} \varphi_i \nabla \varphi_j \\
 M2_{ij} &= \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_i) \cap \text{sop}(\varphi_j)}} \int_K \gamma \varphi_i \varphi_j.
 \end{aligned}$$

Son matrices de orden $N_h \times N_h$, con lo cual $A1 = M + E + M2 \in \mathcal{M}_{N_h \times N_h}$. Computacionalmente se utilizó una función en Matlab con outputs M, E, M2 e input vert, es decir:

```
function [M,E,M2] = stima3(verte).
```

Para el cálculo de la matriz M se utiliza el siguiente código de [9].

```

function M = stima3(vert)
d = size(vert,2);
G = [ones(1,d+1);vertices']\[zeros(1,d);eye(d)];
M = det([ones(1,d+1);vertices'])*G*G'/prod(1:d);

```

De manera análoga adaptamos el código para calcular E y M2.

Para el cálculo de la matriz de rigidez del esquema EAFEM se sigue de manera similar al caso de Galerkin, pero teniendo en cuenta que en este caso la forma bilineal está determinada por $a_h(\cdot, \cdot)$ visto en la Sección 3.2 y para el cálculo del vector de carga se procede de la siguiente manera:

$$\begin{aligned}
 F(\varphi_j) &= \int_{\Omega} f \varphi_j + \int_{\Gamma_D} g \varphi_j \\
 &= \int_{\text{sop}(\varphi_j) \cap \Omega} f \varphi_j + \int_{\text{sop}(\varphi_j) \cap \Gamma_D} g \varphi_j.
 \end{aligned}$$

Dado que, $\text{sop}(\varphi_j) \cap \Omega = \text{sop}(\varphi)$ y $\text{sop}(\varphi_j) \cap \Gamma_D = \emptyset$ se infiere,

$$F(\varphi_j) = \int_{\text{sop}(\varphi_j)} f \varphi_j = \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_j)}} \int_K f \varphi_j.$$

Luego, aplicando la regla de integración numérica del punto medio (en este caso baricentro) se aproxima la integral anterior y se obtiene:

$$F(\varphi_j) = \sum_{\substack{K \in \mathcal{T}_h \\ K \in \text{sop}(\varphi_j)}} \int_K f \varphi_j \approx \sum_{l=1}^{N_{\text{sop}}} |K_l| f(x_{\text{bar}}) \varphi_j(x_{\text{bar}}),$$

donde, N_{sop} = cantidad de elementos que pertenecen al $\text{sop}(\varphi_j)$. Luego teniendo en cuenta la regularidad de la triangulación (elementos equivalentes) resulta:

$$F(\varphi_j) = N_{\text{sop}} |K| f(x_{\text{bar}}) \varphi_j(x_{\text{bar}}).$$

Por la propiedades de linealidad de la funciones bases, se obtiene $\varphi_j(x_{\text{bar}}) = \frac{1}{3}$, lo que implica:

$$F(\varphi_j) = \frac{N_{\text{sop}} |K| f(x_{\text{bar}})}{3}.$$

Por otra parte, para un elemento $|K|$ con coordenadas $x_i = (y_i, z_i)$, se cumple

$$2|K| = \det \begin{pmatrix} y_2 - y_1 & y_3 - y_1 \\ z_2 - z_1 & z_3 - z_1 \end{pmatrix}$$

Esto implica que:

$$F(\varphi_j) = \det \begin{pmatrix} y_2 - y_1 & y_3 - y_1 \\ z_2 - z_1 & z_3 - z_1 \end{pmatrix} \frac{N_{sop}}{6} f(x_{bar}).$$

La manera computacional de adaptar el vector de carga es la siguiente:

```
% Volume Forces
for j = 1:size(ele,1)
    b(ele(j,:)) = b(ele(j,:)) + ...
    det([1,1,1; vert(ele(j,:),:)]') * ...
    F(sum(vert(ele(j,:),:))/3)/6;
end
```

En lo que sigue, se muestra la gráfica de la solución exacta y la solución calculada con los métodos de Galerkin y EAFEM en la Figura 3.4 y la Figura 3.5 respectivamente, para nueve grados de libertad. Además se puede observar que la solución calculada con el método EAFEM se acerca mucho mejor a la solución, mientras que la solución calculada con el método de Galerkin tiene oscilaciones.

Finalmente, para calcular las soluciones numéricas o aproximadas de los esquemas discretos de Galerkin y esquema EAFEM con mayor número de grados de libertad se procede de forma similar que el caso anterior. Recordemos que, la idea de presentar este ejemplo es hacer una comparación de los gráficos entre el esquema EAFEM y esquema de Galerkin. Para ello en ambos métodos utilizamos un número inicial de grados de libertad $N = 9$ en una malla uniforme donde el diámetro de cada triángulo es de $h = 0,5$, y el número máximo de grados de libertad es de $N = 16641$ para una malla también uniforme de $h = 0,0078$.

En la Figura 3.6, se muestra la solución calculada con el método de Galerkin en la primera columna, la solución calculada con el esquema EAFEM se muestra en la columna central, y la

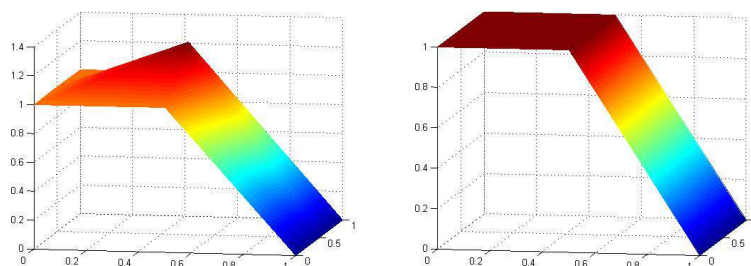


Figura 3.4: Solución calculada con el método clásico (lado izquierdo) utilizando nueve grados de libertad y solución exacta (lado derecho)

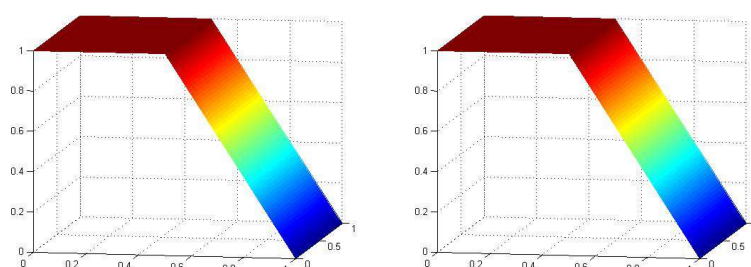


Figura 3.5: Solución calculada con el método EAFEM (lado izquierdo) utilizando 9 grados de libertad y solución exacta (lado derecho)

solución exacta se muestra en la tercera columna. Para calcular estas soluciones se utilizaron $N = 9$ grados de libertad en la primera fila, para la segunda fila 25 grados de libertad y 81 grados de libertad para la tercera fila.

En la Figura 3.7, se muestra la solución calculada con el método de Galerkin en la primera columna, la solución calculada con el esquema EAFEM se muestra en la columna central, y la solución exacta se muestra en la tercera columna. Para el cálculo de estas soluciones se utilizaron 289 grados, 1089 y 4225 grados de libertad para la primera, segunda y tercera fila respectivamente.

Con estos gráficos se puede observar que las soluciones calculadas con el método de Galerkin

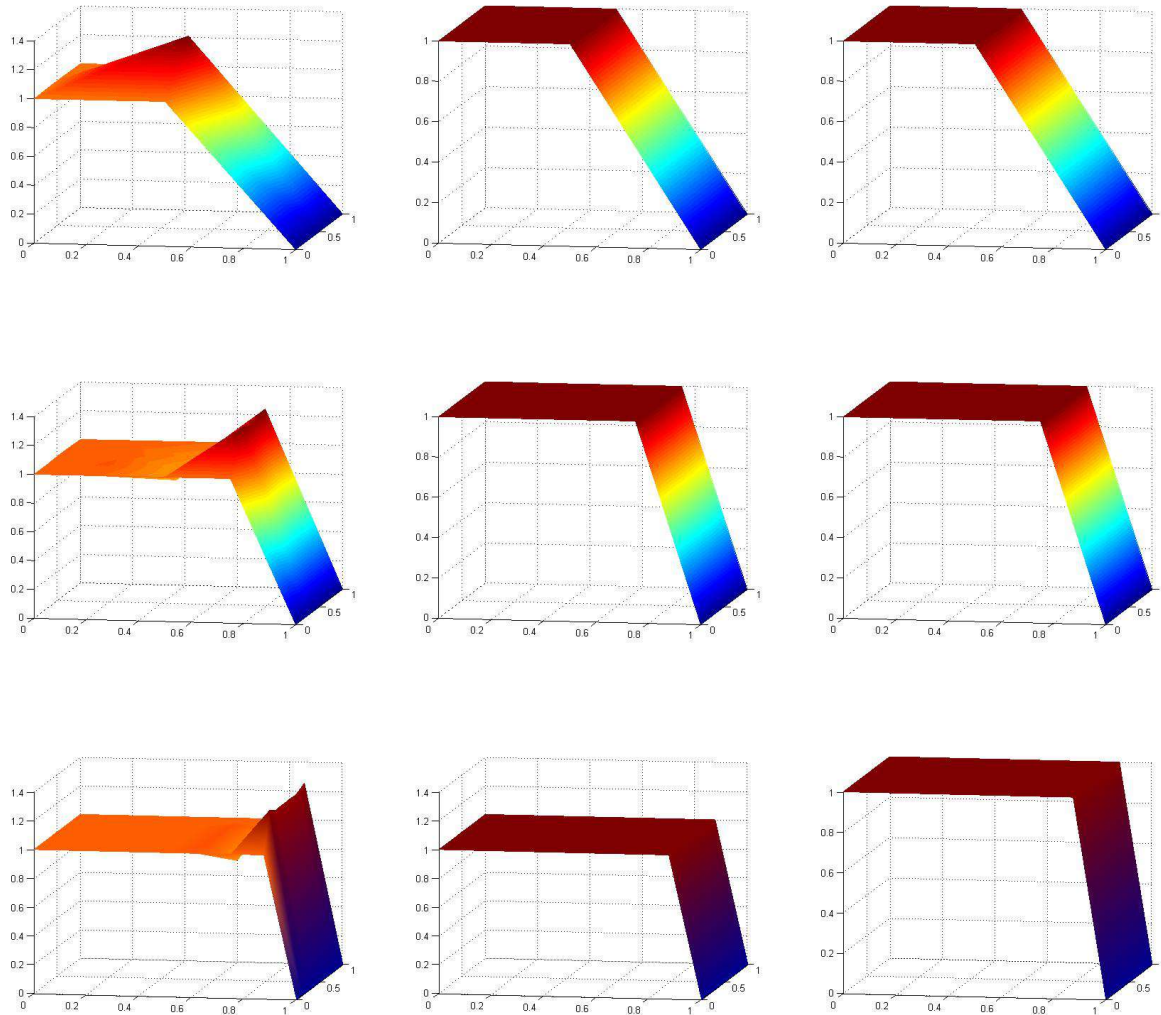


Figura 3.6: Solución calculada con el método de Galerkin (columna del lado izquierdo), solución calculada con el método EAFEM (columna del centro), solución exacta (columna del lado derecho). Malla $h = 0,5$ (primera fila), malla $h = 0,25$ (segunda fila) y malla $h = 0,125$ (tercera fila).

oscilan por que la medida de la malla no es lo suficientemente pequeña, mientras que la solución calculada con el método EAFEM se acerca mucho mejor a la solución exacta.

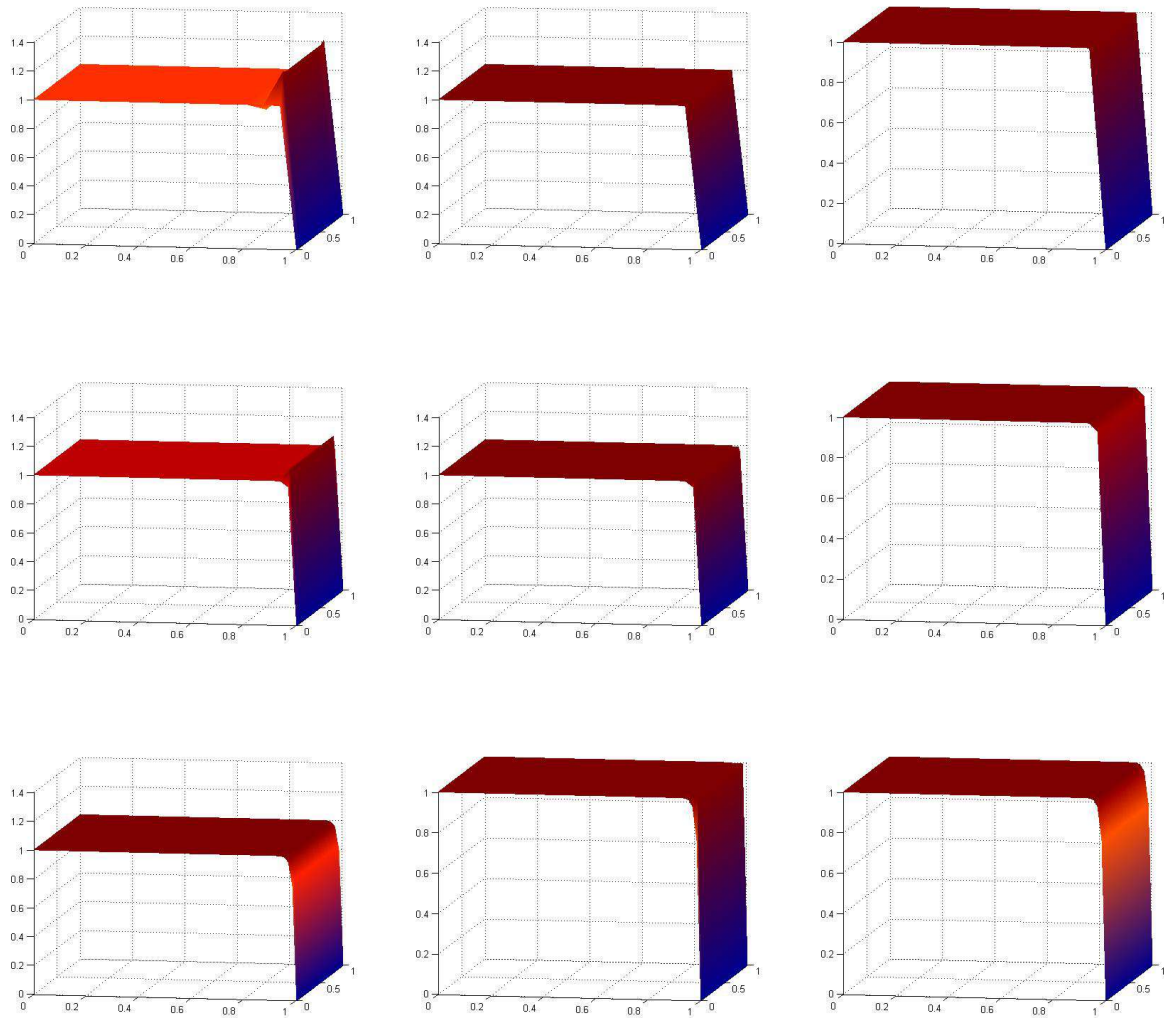


Figura 3.7: Solución calculada con el método de Galerkin (columna del lado izquierdo), solución calculada con el método EAFEM (columna del centro), solución exacta (columna del lado derecho). Malla $h = 0,0625$ (primera fila), malla $h = 0,03125$ (segunda fila) y malla $h = 0,015625$ (tercera fila).

ANEXOS

En esta parte del trabajo se presenta la idea del esquema del código computacional utilizado para el cálculo de las soluciones aproximadas de la ecuación de convección–difusión–reacción del tipo estacionario mediante el método de Galerkin y el método EAFEM, tomando como base la implementación numérica de [9] y [10], que es un paquete de software de de Matlab.

a.- Para el esquema de Galerkin, en un archivo `esqG.m` se añade los siguientes pasos

- 1.- **Partición del dominio:** Se divide el dominio en tantos grados de libertad (vert) y elementos (ele), teniendo en cuenta que la medida de la malla es h (diámetro de cada elemento del mallado)

```
[vert,ele] = squaremesh([0,1,0,1],0.h);  
dirichlet = setboundary(vert,ele,'Dirichlet');  
allEdge = [ele(:, [2,3]); ele(:, [3,1]); ele(:, [1,2])];  
Dir_vec = allEdge((dirichlet(:) == 1),:);  
boundary=unique(Dir_vec(:));  
FreeNodes=setdiff(1:size(vert,1),unique(dirichlet));
```

- 2.- **Término difusivo:** Se llama ε al término difusivo y en el código matlab se escribió como $\varepsilon=v-\text{eps}$.
- 3.- **Ensamble de la matriz de rigidez A_1 :** Se construye una matriz nula con dimensión $N_h \times N_h$. Considerando que, para el caso de Galerkin la forma bilineal es $a(\cdot, \cdot)$

y hacemos llamado de la function `[M,C,E,M2] = stima3(vertices)`, que sirve para integrar las funciones bases sobre las elementos de la triangulación.

```

%%%% ASSEMBLY

row=zeros(9*size(ele,1),1);
col=zeros(9*size(ele,1),1);
valM=zeros(9*size(ele,1),1);
valE=zeros(9*size(ele,1),1);
valM2=zeros(9*size(ele,1),1);
for j = 1:size(ele,1)
    [MT, ET, M2T]=stima3(vert(ele(j,:),:));
    valrow=ones(3,1)*ele(j,:);
    row(9*j-8:9*j,1)=reshape(valrow,1,9);
    col(9*j-8:9*j,1)=reshape(valrow',1,9);
    valM(9*j-8:9*j,1)=reshape(MT,1,9);
    valE(9*j-8:9*j,1)=reshape(ET,1,9);
    valM2(9*j-8:9*j,1)=reshape(M2T,1,9);
end

A=sparse(row,col,valM,size(vert,1),size(vert,1));
D=sparse(row,col,valE,size(vert,1),size(vert,1));
S=sparse(row,col,valM2,size(veret,1),size(vert,1));
A1=v-eps*A+D'+S;

```

4.- **Ensamble del vector de carga b:** Para ello se construye un vector nulo de dimensión $N_h \times 1$ y luego utilizamos el código computacional dado en la Subsección 3.3.1.

5.- **Condiciones de Dirichlet:** Se implementa las condiciones de Dirichlet de la siguiente forma:

```

% Dirichlet conditions
u = sparse(size(vert,1),1);
u(unique(dirichlet))=u_d3(vert(unique(dirichlet),:),v-eps);
b = b - A1 * u;

```

6.- **Solución computacional:** Para calcular la solución aproximada se multiplica el vector de carga por la inversa de la matriz de rigidez.

```
% Computation of the solution
u(FreeNodes) = A1(FreeNodes,FreeNodes) \ b(FreeNodes);
```

b.- Para el esquema EAFEM se sigue de forma similar al caso de Galerkin, la única variación que sufre el código computacional esta en el paso 3, dado que el ensamble de la matriz de rigidez se hace con el siguiente programa:

- **Ensamble de la matriz de rigidez A1:** Para ello se construye una matriz nula con dimensión $N_h \times N_h$ teniendo en cuenta que para el caso del esquema EAFEM la forma bilineal es $a_h(\cdot, \cdot)$ y se hace llamado de

```
function [Mgrad, Mu]=mlocal_quad(T,eps1),
```

para integrar las funciones bases sobre las elementos de la triangulación tal y como se detalla a continuación.

```
%%%% ASSEMBLY
row=zeros(9*size(ele,1),1);
col=zeros(9*size(ele,1),1);
valA=zeros(9*size(ele,1),1);
valB=zeros(9*size(ele,1),1);
for i=1:size(ele,1)
    [AT, BT]=mlocal_quad(verte(ele(i,:),:),v-eps);
    valrow=ones(3,1)*ele(i,:);
    row(9*i-8:9*i,1)=reshape(valrow,1,9);
    col(9*i-8:9*i,1)=reshape(valrow',1,9);
    valA(9*i-8:9*i,1)=reshape(AT,1,9);
    valB(9*i-8:9*i,1)=reshape(BT,1,9);
end
A=sparse(row,col,valA,size(verte,1),size(verte,1));
B=sparse(row,col,valB,size(verte,1),size(verte,1));
A1=A+B;
```


Capítulo 4

Conclusiones

- 1.- Se concluyó existencia y unicidad de la solución aproximada del esquema discreto de Galerkin para la ecuación de convección–difusión–reacción del tipo estacionario utilizando el Lema de Lax–Milgram.
- 2.- Se corroboró existencia y unicidad de la solución aproximada del esquema EAFEM para la ecuación de convección–difusión–reacción del tipo estacionario utilizando el Lema de Lax–Milgram y generalizando la idea propuesta de [5].
- 3.- Se formuló el problema modelo como un esquema de tal manera que se aplicó el Teorema de Lax–Milgram.
- 4.- Se concluyó existencia y unicidad de la solución del problema continuo utilizando el Teorema de Lax–Milgram.
- 5.- Se corroboró la estimación del error entre la solución exacta y la solución aproximada del esquema EAFEM, siguiendo a [5]
- 6.- Se desarrolló un ejemplo numérico que ilustra el funcionamiento de los esquemas de Galerkin y esquema EAFEM.
- 7.- Mediante un ejemplo numérico se comparó el método de Galerkin y el esquema EAFEM dando como resultado que el esquema EAFEM es más eficiente que el esquema de Ga-

lerkin, es decir la solución calculada mediante el esquema EAFEM se aproxima mucho mejor a la solución exacta, mientras que la solución calculada utilizando el esquema de Galerkin tiende a oscilar si la medida de la malla no es suficientemente pequeña.

Capítulo 5

Sugerencias

- 1.- Analizar existencia y unicidad de la solución del problema estacionario sin suponer $\gamma_0 > \frac{\|\mathbf{b}\|_{\infty, \Omega}^2}{2\epsilon_0}$.
- 2.- Desarrollar ejemplos numéricos para comprobar la linealidad del orden de convergencia con respecto a la norma $H^1(\Omega)$.
- 3.- Analizar y detallar las estimaciones apriori del esquema EAFEM propuesto por Xu Zikatanov para \mathbb{R}^3 .
- 4.- Desarrollar un código computacional que sea capaz de testear ejemplos numéricos para \mathbb{R}^3 .

Bibliografía

- [1] BRAMBLE, J.H. & HUBBARD B.E. (1964). *New monotone type approximations for elliptic problems*. Mathematics of Computation, volume 18, number 87, p. 349-367.
- [2] BRAMBLE, J.H. & HUBBARD B.E. (1964). *On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type*. Mathematics and Physics, volume 51, p. 117-132.
- [3] GATICA, G. (2014). *A simple Introducción to the Mixed Element Method. Theory and Applications*. Primera edición, SpringerBriefs in Mathematics.
- [4] GILBARD, D. & TRUDINGER, N. (1983). *Elliptic Partial Differential Equations of Second Order*. First edition, Springer-Verlag, volume 224.
- [5] XU, J. & ZIKATANOV, L. (1999). *A monotone finite element scheme for convection-diffusion equations*. Journal of Computational and Applied Mathematics, volume 68, number 228, p. 1429-1446.
- [6] XU, J. (1996). *Two-grid discretization techniques for linear and nonlinear pdes**. SIAM Journal on Numerical Analysis, volume 33, number 5, p. 1759-1777.
- [7] MARKOWICH, P. & ZLAMAL, M. (1988). *Inverse-average-type finite element discretizations of self-adjoint second-order elliptic problems*. Mathematics of Computation, volume 51, number 184, p. 431-449.

- [8] BREZZI, F., MARINI, L. & PIETRA, P. (1989). *2-dimensional exponential fitting and applications to drift-diffusion models*. SIAM Journal on Numerical Analysis, volume 26, number 6, p. 1342-1355.
- [9] JOCHEN ALBERTY, CARSTEN CARSTENSEN & STEFAN A. FUNKEN (1998). *Remarks Around 50 Lines Of Matlab: Short Finite Element Implementation*. Numerical Algorithms, volume 20, p. 117-137.
- [10] CHEN, L, 2009, *iFEM: An Integrated Finite Element Methods Package in MATLAB*. Technical Report. University of California at Irvine.