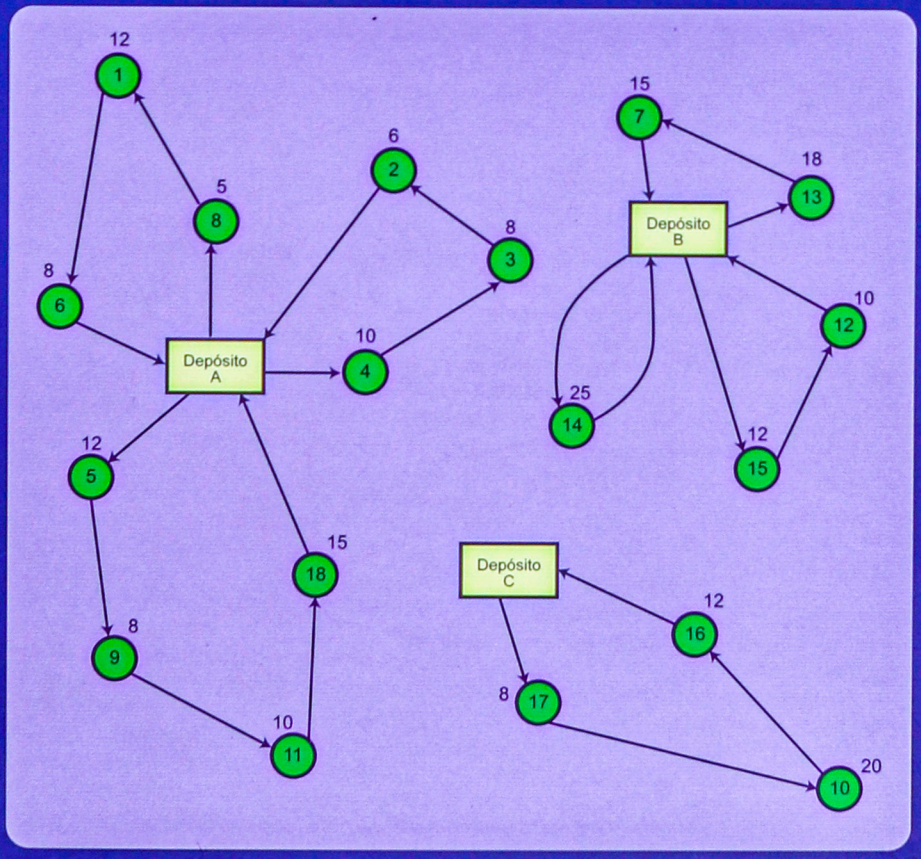
ISSN: 1813-3894

REVCIUNI

VOLUMEN 23 NÚMERO 1 ENERO-DICIEMBRE 2020

Revista de la Facultad de Ciencias de la UNI - REVCIUNI

Publicada por la Unidad de Investigación de la Facultad de Ciencias de la Universidad Nacional de Ingeniería



UNIVERSIDAD NACIONAL DE INGENIERÍA LIMA-PERÚ

AUTORIDADES UNIVERSITARIAS

RECTOR

Dr. Pablo Alfonso Lopez Chau Nava

VICERRECTOR ACADÉMICO

Dra. Shirley Emperatriz Chilet Cama

VICERRECTOR DE INVESTIGACIÓN

Dr. Arturo Fernando Talledo Coronado

DECANO DE LA FACULTAD DE CIENCIAS

Dr. Pedro Canales García

Carátula: *Una solución factible para el problema propuesto en la figura 1. Esta solución usa 7 vehículos y tiene una longitud total de recorrido de 59.5 kilómetros y en consecuencia un costo total igual a 4760 soles.*

Revista de la Facultad de Ciencias de la UNI – REVCIUNI

Volumen 23, Número 1, enero-diciembre 2020
ISSN: 1813-3894

Publicada por la Unidad de Investigación de la Facultad de Ciencias
de la Universidad Nacional de Ingeniería

La revista se distribuye en la Facultad de Ciencias – UNI
Av. Túpac Amaru 210 - Rimac
Lima - Perú
Página web: http://fc.uni.edu.pe/revciuni
E-mail: investigacionfc@uni.edu.pe
Frecuencia de publicación: Anual

HECHO EL DEPÓSITO LEGAL EN LA
BIBLIOTECA NACIONAL DEL PERÚ N° 1813-3894
Revista indexada en el LATINDEX

Impreso: *imprenta FABET e.i.r.l.*
correo: *fabeteirl@yahoo.com*
móvil: *998 434 136*

EDITORIAL

Es muy grato para nuestra Facultad hacer llegar a la comunidad científica y en general a toda nuestra sociedad el Volumen Nro. 23 de nuestra revista REVCIUNI, en la cual nuestros investigadores tienen una oportunidad para hacer públicos los avances de sus investigaciones y así contribuir con la difusión del conocimiento, el cual debe estar decididamente orientado al servicio de la sociedad, y de este modo se pueda lograr el progreso de la nación. La tarea fundamental del quehacer universitario es la investigación, y en la Facultad de Ciencias tiene prioridad al igual que la enseñanza de calidad. Los trabajos que se presentan son en su mayoría producto del trabajo de tesis de nuestros alumnos así como de Proyectos de Investigación que se desarrollan en nuestra Facultad. Esperamos seguir contando con la entusiasta participación de los miembros de nuestra comunidad académica, lo cual permitirá seguir incrementando el número de artículos que se publican en nuestra revista.

Dr. Pedro Canales García
Decano
Facultad de Ciencias
Universidad Nacional de Ingeniería

Sucesión Espectral de Grothendieck en Homología de Grupos

Felipe Clímaco Ccolque Taipe
Instituto de Matemática y Ciencias Afines. Facultad de Ciencias.
Universidad Nacional de Ingeniería;
ccolque@imca.edu.pe

Recibido el 3 de Febrero de 2020; aceptado 22 de Setiembre de 2020

En este artículo se demuestra la existencia de la sucesión espectral homológica de Grothendieck. Dados K un grupo y N un subgrupo normal de K se establece una relación entre los grupos de homología de K, N y K/N, utilizando estos resultados.

Palabras Claves: Grupos de homología, K- módulos, sucesión espectral homológica de Grothendieck, lema de herradura, resolución proyectiva de un objeto en una categoría abeliana.

In this article, the existence of the homological version of Grothendieck spectral sequence is demostrated. Given K a group and N a normal subgroup of K, a relationship between the groups of homology of K, N and K/N is established, using these results.

Keywords: Homology groups, K-modules, Grothendiek homological spectral sequence, horseshoe lemma, projective resolution of an object in an abelian category.

1 Introducción

Las sucesiones espectrales fueron introducidas por Jean Leray en 1946. Las sucesiones espectrales son herramientas fundamentales en topología algebraica, geometría algebraica [1], álgebra homológica, teoría de números, variedades complejas [2] y $K-$teoría.
Grothendieck introdujo una sucesión espectral que relaciona los funtores derivados de un funtor compuesto de dos funtores, y los funtores derivados de los factores. En el contexto de cohomología este resultado es probado en [3], [4], [5] y [6].
En este artículo, modificando la prueba dada en [3] y considerando [7], se consigue la prueba del resultado en el contexto homológico.

Por otro lado en [3], se probó el resultado siguiente:
Teorema 9.5 (Lyndon-Hochschild-Serre) Dada la sucesión exacta corta de grupos
$N \xrightarrow{\ i\ } K \xrightarrow{\ p\ } Q$ y dado un K-módulo A, existe una sucesión espectral $E = \{E_n(A)\}$ tal que

$$E_1^{p,q} = H^p\left(Q, H^{q-p}(N, A)\right) \Rightarrow H^q(K, A)\ . \qquad (1)$$

Es decir, la sucesión espectral E converge finitamente al grupo graduado asociado al grupo de cohomología $\{H^q(K, A)\}$, filtrado adecuadamente.
Para la prueba se utilizó como herramienta la sucesión espectral cohomológica de Grothendieck , la definición de grupo de cohomología de grupos y propiedades de anillos de grupo con coeficientes enteros.
Teniendo en cuenta la sugerencia dada en [3], en observación ii), después de la prueba de theorem 9.5, se obtendrá la prueba de la versión homológica de la sucesión espectral de Lyndon-Hochschild-Serre.

Este artículo está organizado como sigue:
En la sección 2 se da una revisión de complejos dobles de cadenas. En la sección 3 se prueba la existencia de la sucesión espectral homológica de Grothendieck. En la sección 4 se establece la relación de los tres grupos de homologías de un grupo, de un subgrupo normal y grupo cociente correspondiente.

2 Complejos Dobles de Cadenas y Filtraciones de sus Complejos Totales

La existencia de la sucesión espectral homológica de Grothendieck y el problema de convergencia finita correspondiente están relacionados con el estudio de complejos dobles de cadenas y filtraciones de los complejos totales respectivos. El propósito de esta parte será mostrar resultados que permiten hallar los dos primeros términos de la sucesión espectral que se construye a partir de un complejo doble de cadenas con una filtración de su complejo total, y conseguir la convergencia finita en el caso en que el complejo doble de cadenas sea positivo.

Sea $(B, \partial', \partial'')$ un complejo doble de cadenas sobre alguna categoría abeliana $\mathfrak{A}$ como el dado en [3, p 167]. Se tiene el diagrama siguiente anticonmutativo en $\mathfrak{A}$

$$\begin{array}{ccc} B_{r-1,s} & \xleftarrow{\ \partial'\ } & B_{rs} \\ \downarrow{\scriptstyle \partial''} & & \downarrow{\scriptstyle \partial''} \\ B_{r-1,s-1} & \xleftarrow{\ \partial'\ } & B_{r,s-1} \end{array} \qquad (2)$$

$\partial''\partial' + \partial'\partial'' = 0$ para cada $r, s \in \mathbb{Z}$; es conveniente reemplazar (2) por un diagrama conmutativo (4) y viceversa; esto se logra haciendo

$$d' = \partial', \ d'' = (-1)^r \partial'' \quad \text{sobre} \quad B_{rs} \tag{3}$$

Se sabe que el complejo total de B se define como $Tot\, B = \{(Tot\, B)_n\}$, $n \in \mathbb{Z}$
donde $(Tot\, B)_n = \bigoplus_{r+s=n} B_{rs}$. Es conveniente subrayar que $Tot\, B$ es un complejo de cadenas si el diferencial en $Tot\, B$ es dado por $\partial = \partial' + \partial'' : Tot\, B \to Tot\, B$.
Si se considera el diagrama

$$\begin{array}{ccc} B_{r-1,s} & \xleftarrow{\ d'\ } & B_{rs} \\ \downarrow{\scriptstyle d''} & & \downarrow{\scriptstyle d''} \\ B_{r-1,s-1} & \xleftarrow{\ d'\ } & B_{r,s-1} \end{array} \tag{4}$$

entonces d', d'' se llaman el diferencial horizontal y vertical de B, respectivamente.

El complejo $Tot\, B$ puede ahora filtrarse en los siguientes dos modos de manera natural:

$${}_1F_p(Tot\, B)_n = \bigoplus_{\substack{r+s=n \\ r \le p}} B_{r,s} \tag{5}$$

$${}_2F_p(Tot\, B)_n = \bigoplus_{\substack{r+s=n \\ s \le p}} B_{r,s} \tag{6}$$

Se referirá a la filtración (5) como la PRIMERA FILTRACIÓN de $Tot\, B$, y la filtración (6) como la SEGUNDA FILTRACIÓN de $Tot\, B$. Con estas filtraciones se obtienen dos sucesiones espectrales denotadas por ${}_1E$ y ${}_2E$.

Las pruebas de Proposiciones 2.1 y 2.3 se pueden encontrar en [3].

Proposición 2.1. *Para la primera sucesión espectral asociada a la filtración (5) se tiene*

$${}_1E_0^{p,q} = H_{q-p}(B_{p,*}; \partial''), \ {}_1E_1^{p,q} = H_p(H_{q-p}(B, \partial''), \partial') \tag{7}$$

Para la segunda sucesión espectral asociada a la filtración (6) se tiene

$${}_2E_0^{p,q} = H_{q-p}(B_{*,p}; \partial'), \quad {}_2E_1^{p,q} = H_p(H_{q-p}(B, \partial'), \partial'') \tag{8}$$

Definición 2.2. *Se dice que el complejo doble B (Figura 1) es positivo si existe $n_0 \in \mathbb{Z}$ tal que (el desplazamiento hacia la derecha se considera positivo)*

$$B_{r,s} = 0 \quad si \quad r < n_0 \quad o \quad s < n_0 \tag{9}$$

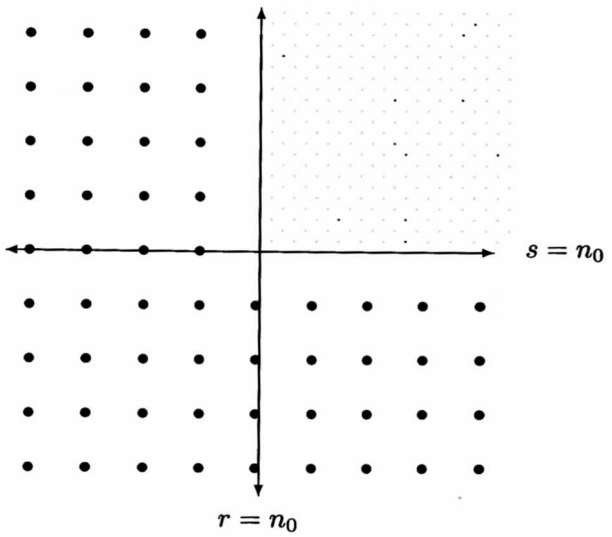


Figura 1. Cada punto marcado de la región no sombreada es $B_{rs} = 0$.

Proposición 2.3. *Si B es un complejo de cadenas doble positivo, entonces ambas la primera y la segunda sucesión espectral ${}_1E$ y ${}_2E$, de las cuales dos primeros términos se dan en (7) y (8), respectivamente; convergen finitamente al correspondiente objeto graduado asociado a la homología $\{H_n(Tot\, B)\}$, la cual es adecuada y finitamente filtrada.*

3 Sucesión Espectral Homológica de Grothendieck

En esta sección se enuncia y se demuestra el teorema de sucesión espectral homológica de Grothendieck que relaciona a los funtores derivados izquierdos.

Existencia y Convergencia de la Sucesión Espectral de Grothendieck

La construcción de la sucesión espectral homológica de Grothendieck se hará usando un complejo doble de cadenas J de objetos proyectivos y este complejo para su construcción requiere de una herramienta llamada LEMA DE HERRADURA. En lo que sigue se abordará el lema mencionado y otros resultados previos que son necesarios para demostrar la existencia de la sucesión espectral homológica de Grothendieck.

***Lema* 3.1.** *Sean P_1 y P_2 objetos proyectivos en una categoría abeliana $\mathfrak{A}$, entonces $P_1 \oplus P_2$ es proyectivo.*

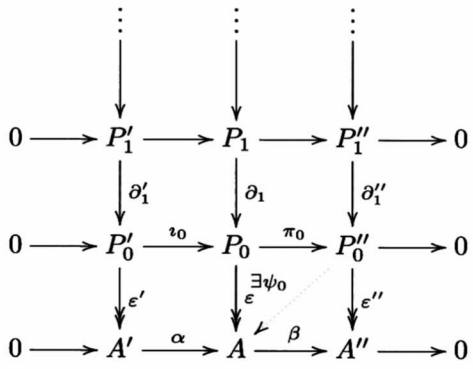
Prueba.- Sea $\alpha : P_1 \oplus P_2 \longrightarrow B$ un morfismo y $\varepsilon : A \twoheadrightarrow B$ un epimorfismo, entonces existe un morfismo $\beta : P_1 \oplus P_2 \longrightarrow A$ tal que $\alpha = \varepsilon \circ \beta$.

En efecto, sean $\alpha_1 : P_1 \longrightarrow B$, $\alpha_2 : P_2 \longrightarrow B$ morfismos. Como P_1 y P_2 son objetos proyectivos, existen morfismos $\beta_1 : P_1 \longrightarrow A$, $\beta_2 : P_2 \longrightarrow A$ tales que $\alpha_1 = \varepsilon \circ \beta_1$, $\alpha_2 = \varepsilon \circ \beta_2$.

$$\begin{aligned} \text{Definiendo } \alpha(a,b) &= \varepsilon \circ \beta_1(a) + \varepsilon \circ \beta_2(b) \\ &= \varepsilon \circ (\beta_1, \beta_2)(a,b), \end{aligned}$$

existe un morfismo $\beta = (\beta_1, \beta_2)$ tal que $\alpha = \varepsilon \circ \beta$ ■

***Lema* 3.2** (Lema de Herradura). *Sea $\mathfrak{A}$ una categoría abeliana y $0 \to A' \to A \to A'' \to 0$ una sucesión exacta corta en $\mathfrak{A}$. Si $\mathcal{P}' : 0 \leftarrow P_0' \leftarrow P_1' \leftarrow \cdots$ y $\mathcal{P}'' : 0 \leftarrow P_0'' \leftarrow P_1'' \leftarrow \cdots$ son resoluciones proyectivas de A' y A'', respectivamente; entonces existe una resolución proyectiva $\mathcal{P}$ de A tal que la sucesión de complejos de cadenas $0 \to \mathcal{P}' \to \mathcal{P} \to \mathcal{P}'' \to 0$ es exacta y tal que el diagrama siguiente es conmutativo*



(10)

Prueba.- Sea $P_0 = P_0' \oplus P_0''$. Como P_0' y P_0'' son proyectivos, por Lema 3.1 P_0 es proyectivo.
Se define $\varepsilon : P_0 \to A$ por $\varepsilon(a,b) = \alpha\varepsilon' a + \psi_0 b$. También se definen los morfismos $\imath_0 : P_0' \to P_0$ por $\imath_0(a) = (a,0)$ y $\pi_0 : P_0 \to P_0''$ por $\pi_0(a,b) = b$.
Se prueba que los dos cuadriláteros inferiores del diagrama (10) son conmutativos, verificando que $\varepsilon\imath_0 = \alpha\varepsilon'$ y $\varepsilon''\pi_0 = \beta\varepsilon$, como sigue:
Sea $a \in P_0'$, entonces $\varepsilon\imath_0(a) = \varepsilon(a,0) = \alpha\varepsilon' a$.

Por otro lado, para $(a,b) \in P_0 = P_0' \oplus P_0''$, $\varepsilon''\pi_0(a,b) = \varepsilon''(b) = \beta\psi_0 b$ pues P_0'' es proyectivo y

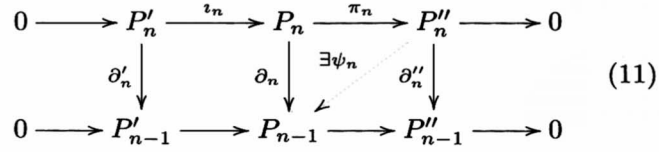
$$\begin{aligned} \beta\varepsilon(a,b) &= \beta(\alpha\varepsilon' a + \psi_0 b) \\ &= \beta\alpha\varepsilon' a + \beta\psi_0 b, \text{como } \beta\alpha = 0 : \\ &= \beta\psi_0 b \,. \end{aligned}$$

Así, $\varepsilon''\pi_0 = \beta\varepsilon$.
Puesto que $Im(\imath_0) = P_0' \times \{0\} = Ker(\pi_0)$, resulta que $0 \to P_0' \xrightarrow{\imath_0} P_0 \xrightarrow{\pi_0} P_0'' \to 0$ es sucesión exacta de proyectivos. De la conmutatividad de los diagramas con filas exactas cortas, por ser ε' y ε'' epimorfismos, se sigue que ε es un epimorfismo.

HIPÓTESIS INDUCTIVA: Suponemos que existe sucesión exacta corta de proyectivos
$0 \longrightarrow P_m' \xrightarrow{\imath_m} P_m \xrightarrow{\pi_m} P_m'' \longrightarrow 0$ para $0 \le m < n$, donde $P_m = P_m' \oplus P_m''$, tal que el diagrama (10) hasta dicha sucesión exacta es conmutativo.
Ponemos $P_n = P_n' \oplus P_n''$. Al igual que arriba construimos la sucesión exacta corta de proyectivos

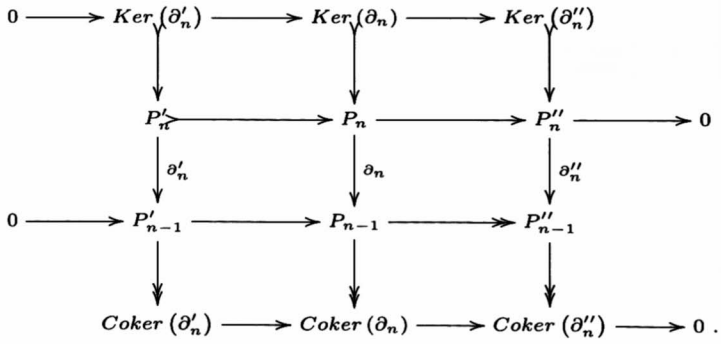
$$0 \to P_n' \to P_n \to P_n'' \to 0$$

tal que el diagrama

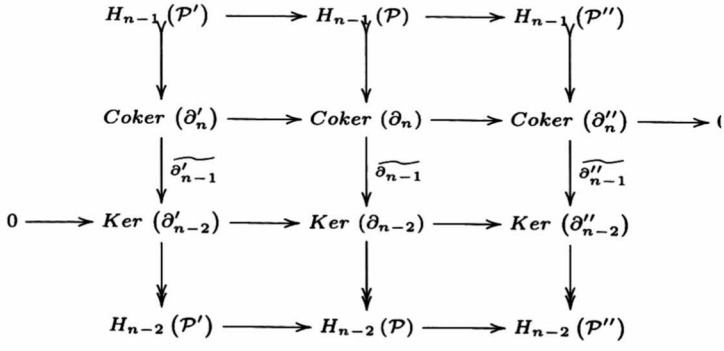


(11)

donde $\partial_n(a,b) = \imath_{n-1}\partial_n' a + \psi_n b$, es conmutativo.
Además $Im(\partial_n) = Ker(\partial_{n-1})$, lo cual se verifica mostrando que $H_{n-1}(\mathcal{P}) = 0$ para $n-1 \ge 0$.
Para ello, se considera el diagrama



Por [3, Lema III.5.1], las sucesiones de las filas superior e inferior son exactas. Según [3, Lema IV.2.2]:



Por [3, Lema III.5.1] $H_{n-1}(\mathcal{P}') \to H_{n-1}(\mathcal{P}) \to H_{n-1}(\mathcal{P}'')$ es exacta. Pero $H_{n-1}(\mathcal{P}') = H_{n-1}(\mathcal{P}'') = 0$,

luego $H_{n-1}(\mathcal{P}) = 0$.
Por el principio de inducción, existe un complejo de cadenas $\mathcal{P} : 0 \leftarrow P_0 \leftarrow P_1 \leftarrow \cdots$, donde cada P_n es proyectivo, $\mathcal{P}$ es acíclico ($H_n(\mathcal{P}) = 0$, $\forall n \geq 1$) y $H_0(\mathcal{P}) \cong A$. Así, $\mathcal{P}$ es una resolución proyectiva de A tal que $0 \to \mathcal{P}' \to \mathcal{P} \to \mathcal{P}'' \to 0$ es sucesión exacta de complejos de cadenas pues $0 \to \mathcal{P}'_n \to \mathcal{P}_n \to \mathcal{P}''_n \to 0$ es exacta para cada $n = 0, 1, \ldots$. Además el diagrama (10) es conmutativo ■

***Lema* 3.3.** *Sea $\mathfrak{C}$ una categoría abeliana con suficientes proyectivos.*
Sean $0 \xleftarrow{\partial_0} F_0 \xleftarrow{\partial_1} F_1 \longleftarrow \cdots \xleftarrow{\partial_r} F_r \xleftarrow{\partial_{r+1}} \cdots$ un complejo de cadenas en $\mathfrak{C}$, $Z_{r+1} = Ker(\partial_{r+1})$ y $B_r = Im(\partial_{r+1})$ para $r = 0, 1, \ldots$; entonces existe una resolución proyectiva de

$$F_0 \longleftarrow B_0 \twoheadleftarrow F_1 \longleftarrow Z_1 \longleftarrow B_1 \twoheadleftarrow F_2 \longleftarrow \cdots \tag{12}$$

expresada como el diagrama conmutativo

$$\begin{array}{ccccccccccc}
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
J_{01} & \longleftarrow & L_{01} & \twoheadleftarrow & J_{11} & \longleftarrow & K_{11} & \longleftarrow & L_{11} & \twoheadleftarrow & J_{21} & \longleftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
J_{00} & \longleftarrow & L_{00} & \twoheadleftarrow & J_{10} & \longleftarrow & K_{10} & \longleftarrow & L_{10} & \twoheadleftarrow & J_{20} & \longleftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
F_0 & \longleftarrow & B_0 & \twoheadleftarrow & F_1 & \longleftarrow & Z_1 & \longleftarrow & B_1 & \twoheadleftarrow & F_2 & \longleftarrow & \cdots
\end{array} \tag{13}$$

donde cada columna es una resolución proyectiva completa del objeto que aparece en su pie y $L_{rs} \twoheadleftarrow J_{r+1,s} \longleftarrow K_{r+1,s}$ es exacta.

Prueba .- La sucesión $F_0/B_0 \twoheadleftarrow F_0 \longleftarrow B_0$ es exacta corta. Como $\mathfrak{C}$ tiene suficientes proyectivos, se puede asumir existen resoluciones proyectivas Q_0 de F_0/B_0, y L_0 de B_0. Por el lema de herradura, se obtiene una resolución proyectiva J_0 de F_0 tal que el diagrama siguiente conmuta

$$\begin{array}{ccccc}
\vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow \\
Q_{01} & \longleftarrow & J_{01} & \longleftarrow & L_{01} \\
\downarrow & & \downarrow & & \downarrow \\
Q_{00} & \twoheadleftarrow & J_{00} & \longleftarrow & L_{00} \\
\downarrow & & \downarrow & & \downarrow \\
F_0/B_0 & \twoheadleftarrow & F_0 & \longleftarrow & B_0
\end{array} \tag{14}$$

Escogiendo las resoluciones proyectivas L_1 de B_1 y Q_1 de Z_1/B_1, se obtiene una resolución proyectiva K_1 de Z_1 tal que el diagrama siguiente conmuta

$$\begin{array}{ccccc}
\vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow \\
Q_{11} & \longleftarrow & K_{11} & \longleftarrow & L_{11} \\
\downarrow & & \downarrow & & \downarrow \\
Q_{10} & \twoheadleftarrow & K_{10} & \longleftarrow & L_{10} \\
\downarrow & & \downarrow & & \downarrow \\
Z_1/B_1 & \twoheadleftarrow & Z_1 & \longleftarrow & B_1
\end{array} \tag{15}$$

Con las resoluciones proyectivas L_0 de B_0, K_1 de Z_1, aplicando Lema de herradura se obtiene la resolución proyectiva J_1 de F_1 tal que el diagrama siguiente conmuta :

$$\begin{array}{ccccc}
\vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow \\
L_{01} & \twoheadleftarrow & J_{11} & \longleftarrow & K_{11} \\
\downarrow & & \downarrow & & \downarrow \\
L_{00} & \twoheadleftarrow & J_{10} & \longleftarrow & K_{10} \\
\downarrow & & \downarrow & & \downarrow \\
B_0 & \twoheadleftarrow & F_1 & \longleftarrow & Z_1
\end{array} \tag{16}$$

Acoplando los diagramas (14), (15) y (16) se obtiene una parte importante del diagrama (13).

Repitiendo el procedimiento anterior se obtiene la resolución proyectiva J_2 de F_2 de tal modo que el diagrama (13) conmuta, donde cada columna es resolución proyectiva completa del objeto que aparece en su pie y que $L_{rs} \twoheadleftarrow J_{r+1,s} \leftarrow K_{r+1,s}$ es exacta ■

Proposición 3.4. *Del diagrama (13) se obtiene el diagrama conmutativo*

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\downarrow & & \downarrow & & \downarrow & & \\
J_{01} & \leftarrow & J_{11} & \leftarrow & J_{21} & \leftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
J_{00} & \leftarrow & J_{10} & \leftarrow & J_{20} & \leftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
F_0 & \leftarrow & F_1 & \leftarrow & F_2 & \leftarrow & \cdots
\end{array} \tag{17}$$

donde cada fila es un complejo de cadenas y la $r-$ésima columna es una resolución proyectiva completa de F_r; $r = 0, 1, \ldots$.

Prueba .- La conmutatividad del diagrama (17) se obtiene de la conmutatividad del diagrama (13).
La fila base del diagrama (17) por hipótesis es un complejo de cadenas. Del hecho que
$L_{rs} \twoheadleftarrow J_{r+1,s} \leftarrow K_{r+1,s}$ es exacta, se sigue que $J_{rs} \leftarrow J_{r+1,s} \leftarrow J_{r+2,s}$ es el morfismo nulo. Así, cada fila del diagrama (17) es un complejo de cadenas. Por Lema 3.3, $J_r = J_{r,*}$ es una resolución proyectiva de F_r para $r = 0, 1, \ldots$ ■

El siguiente complejo doble de cadenas se denotará por D:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\downarrow & & \downarrow & & \downarrow & & \\
C_{01} & \leftarrow & C_{11} & \leftarrow & C_{21} & \leftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
C_{00} & \leftarrow & C_{10} & \leftarrow & C_{20} & \leftarrow & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
A_0 & \leftarrow & A_1 & \leftarrow & A_2 & \leftarrow & \cdots
\end{array} \tag{18}$$

Proposición 3.5. *Si las homologías de las columnas de D se anulan, entonces el complejo total de D tiene homología nula.*

Prueba .- Sea $D_j = {}_1F_j(Tot\ D)$ subcomplejo de $(Tot\ D)$ y C_j columna $j-$ésima de D, entonces $C_j = D_j/D_{j-1}$. Así, la sucesión de complejos de cadenas $D_{j-1} \rightarrowtail D_j \twoheadrightarrow C_j$ es exacta.
Por [3, Teorema IV.2.1], la sucesión larga de homologías siguiente es exacta
$\cdots \to H_q(D_{j-1}) \to H_q(D_j) \to H_q(C_j) \to H_{q-1}(D_{j-1}) \to H_{q-1}(D_j) \to H_{q-1}(C_j) \to \cdots$.
Puesto que $H_q(C_j) = 0$ para todo q, j, y $H_q(D_{-1}) = 0$, se prueba por inducción que $H_q(D_j) = 0$ para todo q, j. Pero $(D_j)_q = (TotD)_q$ para $q < j$, luego $H_q(TotD) = H_q(D_{q+2}) = 0$, como se quería probar ■

Se va a considerar tres categorías abelianas $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ y dos funtores covariantes aditivos
$F : \mathfrak{A} \to \mathfrak{B}$, $G : \mathfrak{B} \to \mathfrak{C}$. Además, se considera que $\mathfrak{A}$ y $\mathfrak{B}$ tienen suficientes proyectivos; esto significa que todos los objetos en $\mathfrak{A}$ y $\mathfrak{B}$ tienen resoluciones proyectivas. Así es posible construir los funtores derivados izquierdos de F, G y $G \circ F$. El teorema siguiente establece una relación entre estos funtores derivados mediante una sucesión espectral.

Definición 3.6. *Se dice que un objeto B de $\mathfrak{B}$ es G-acíclico (izquierdo) si*

$$L_qG(B) = \begin{cases} G(B), & q = 0 \\ 0, & q \geq 1 \end{cases} \tag{19}$$

Teorema 3.7 (Sucesión Espectral Homológica de Grothendieck). *Sean $F : \mathfrak{A} \to \mathfrak{B}$ y $G : \mathfrak{B} \to \mathfrak{C}$ funtores covariantes aditivos de categorías abelianas, donde $\mathfrak{A}$ y $\mathfrak{B}$ tienen suficientes proyectivos. Si para cada objeto proyectivo P de $\mathfrak{A}$ se tiene que $F(P)$ es G-acíclico, entonces para cada objeto A de $\mathfrak{A}$ existe una sucesión espectral $E = \{E^n(A)\}$ tal que*

$$E^1_{pq} = (L_pG)(L_{q-p}F)(A) \Rightarrow L_q(GF)(A) \ . \tag{20}$$

Es decir, la sucesión espectral E converge finitamente al objeto graduado asociado a $\{L_q(GF)(A)\}$, filtrado adecuadamente.

Prueba.- Puesto que $A \in |\mathfrak{A}|$ y $\mathfrak{A}$ tiene suficientes proyectivos, existe una resolución proyectiva $\mathcal{P}$ de A escrita como

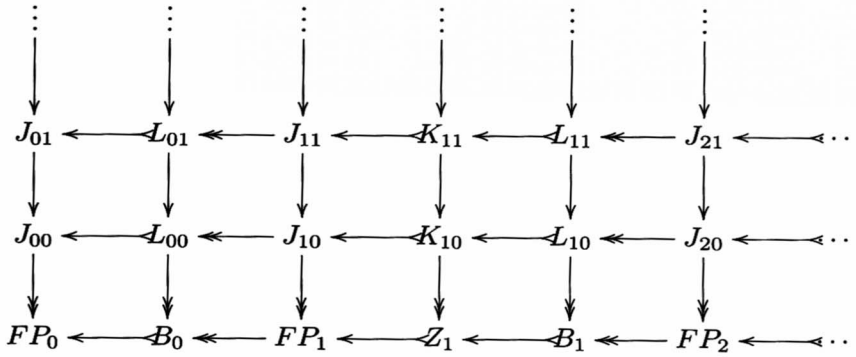
$$\mathcal{P} \ : \ 0 \leftarrow P_0 \leftarrow P_1 \leftarrow \cdots \leftarrow P_n \leftarrow \cdots \tag{21}$$

donde P_n es proyectivo para $n = 0, 1, \ldots$.
Luego

$$F\mathcal{P} \ : 0 \xleftarrow{\partial_0} FP_0 \xleftarrow{\partial_1} FP_1 \leftarrow \cdots \xleftarrow{\partial_{q-p}} FP_{q-p} \xleftarrow{\partial_{q-p+1}} \cdots \tag{22}$$

$$\text{Así, } L_{q-p}F(A) = H_{q-p}(F\mathcal{P}) = \frac{Ker\,(\partial_{q-p})}{Im\,(\partial_{q-p+1})} = \frac{Z_{q-p}}{B_{q-p}} \tag{23}$$

Del Lema 3.3 se obtiene



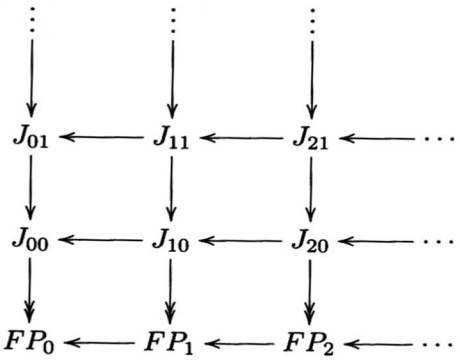
(24)

donde cada columna es una resolución proyectiva completa del objeto que aparece en su pie y la sucesión $L_{rs} \longleftarrow J_{r+1,s} \longleftarrow K_{r+1,s}$ es exacta. Como $\frac{Z_{q-p}}{B_{q-p}} \in |\mathfrak{B}|$ y $\mathfrak{B}$ tiene suficientes proyectivos, este objeto tiene como resolución proyectiva $\frac{K_{q-p}}{L_{q-p}}$.

Por lo tanto,

$$(L_pG)\left(\frac{Z_{q-p}}{B_{q-p}}\right) = H_p\left(G\left(\frac{K_{q-p}}{L_{q-p}}\right)\right)$$
$$= \frac{Ker\left(G\left(\frac{K_{q-p,p}}{L_{q-p,p}}\right) \longrightarrow G\left(\frac{K_{q-p,p-1}}{L_{q-p,p-1}}\right)\right)}{Im\left(G\left(\frac{K_{q-p,p+1}}{L_{q-p,p+1}}\right) \longrightarrow G\left(\frac{K_{q-p,p}}{L_{q-p,p}}\right)\right)} \quad (25)$$

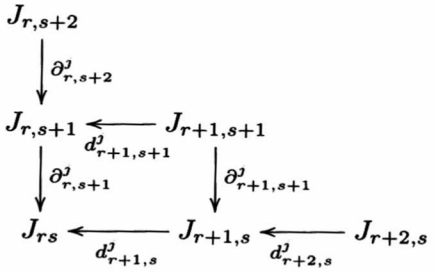
De la Proposición 3.4 obtenemos el diagrama conmutativo



(26)

donde cada fila es un complejo de cadenas y la $r-$ésima columna es una resolución proyectiva completa de FP_r; $r = 0, 1, \ldots$.

Consideramos



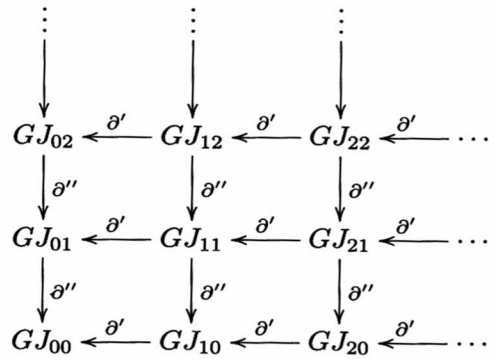
Como G es funtor covariante aditivo, se tiene :

a) $G\left(d^J_{r+1,s}\right) G\left(d^J_{r+2,s}\right) = 0$;

b) $G\left(\partial^j_{r,s+1}\right) G\left(\partial^j_{r,s+2}\right) = 0$;

c) $G\left(\partial^J_{r,s+1}\right) G\left(d^J_{r+1,s+1}\right) = G\left(d^J_{r+1,s}\right) G\left(\partial^J_{r+1,s+1}\right)$.

Aplicando G a la parte J del diagrama (26) y definiendo $\partial' = Gd$ y $\partial'' = (-1)^r G\partial$ morfismos que toman valores en GJ_{rs}; con los items *a)*, *b)* y *c)* se obtiene el complejo doble de cadenas $B = GJ$



(27)

El complejo doble de cadenas GJ es positivo, pues existe $n_0 = 0 \in \mathbb{Z}$ tal que $GJ_{rs} = 0$ si $r < 0 \vee s < 0$. Por Proposición 2.3 existe una sucesión espectral ${}_2E =_2 E(A)$ asociado al complejo de cadenas filtrado $Tot\ GJ$, cuya filtración es dada por

$${}_2F_p\left(Tot\ GJ\right)_n = \bigoplus_{\substack{r+s=n \\ s\leq p}} GJ_{rs}\ ,$$

dicha sucesión converge finitamente al objeto graduado asociado a la homología $\{H_q(Tot\ GJ)\}$ la cual es adecuadamente filtrada. Esto se simboliza por

$${}_2E^1_{pq} \Rightarrow H_q\left(Tot\ GJ\right)\ . \quad (28)$$

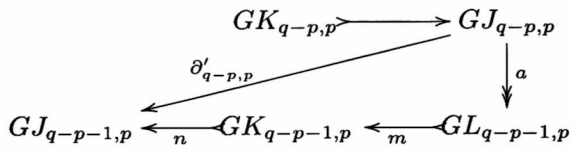
Para finalizar la prueba, falta calcular los dos objetos ${}_2E^1_{pq}$ y $H_q\left(Tot\ GJ\right)$.
Proposición 2.1 da las fórmulas siguientes:

$${}_2E^0_{pq} = H_{q-p}\left(GJ_{*,p}; \partial'\right)\ ,\ {}_2E^1_{pq} = H_p\left(H_{q-p}(GJ, \partial'), \partial''\right)$$

para $B = GJ$, que permiten calcular ${}_2E^1_{pq}$, en los pasos *d)* y *e)*, como sigue :

d)
$$
\begin{aligned}
{}_2E^0_{pq} &= H_{q-p}\left(GJ_{*,p};\partial'\right)\\
&= \frac{Ker\left(\partial'_{q-p,p}\right)}{Im\left(\partial'_{q-p+1,p}\right)}\,. \qquad (29)
\end{aligned}
$$

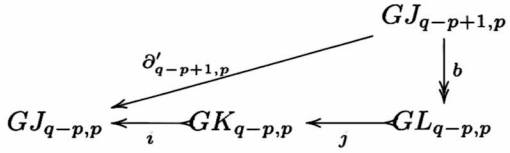
Observando el diagrama (24) se tiene $K_{rs} \longleftarrow L_{rs} \twoheadleftarrow J_{r+1,s} \longleftarrow K_{r+1,s}$.
Como $J_{r+1,s} = L_{rs} \oplus K_{r+1,s}$ y $K_{rs} = Q_{rs} \oplus L_{rs}$ (ver diagramas (15), (16)), aplicando el funtor covariante aditivo G se obtiene
$GK_{rs} \longleftarrow GL_{rs} \twoheadleftarrow GJ_{r+1,s} \longleftarrow GK_{r+1,s}$
Tomando $(r,s) = (q-p-1,p)$ se tiene



luego se deduce que $Ker\partial'_{q-p,p} = GK_{q-p,p}$, pues

$$
\begin{aligned}
& Ker\partial'_{q-p,p}\\
=\ & \{x \in GJ_{q-p,p} \ / \ n \circ m \circ a\,(x) = 0\}\\
=\ & \{x \in GJ_{q-p,p} \ / \ a\,(x) = 0\},\ n \circ m \text{ es monic}\\
=\ & GK_{q-p,p},\ GJ_{q-p,p} = GL_{q-p-1,p} \oplus GK_{q-p,p}\ .
\end{aligned}
$$

Utilizando el diagrama



se obtiene que $Im\left(\partial'_{q-p+1,p}\right) = Im\,(ijb) = GL_{q-p,p}$.

$$\text{Por lo tanto, por (29),}\quad {}_2E^0_{pq} = \frac{GK_{q-p,p}}{GL_{q-p,p}} \qquad (30)$$

Puesto que $GK_{q-p,p} = GQ_{q-p,p} \oplus GL_{q-p,p}$ donde $Q_{q-p,p} = K_{q-p,p}/L_{q-p,p}$, resulta que $\dfrac{G(K_{q-p,p})}{G(L_{q-p,p})} = G\left(\dfrac{K_{q-p,p}}{L_{q-p,p}}\right)$.

$$\text{Luego, } {}_2E^0_{pq} = G\left(\frac{K_{q-p,p}}{L_{q-p,p}}\right)\ . \qquad (31)$$

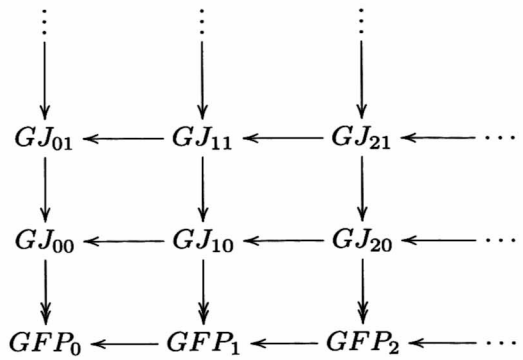
e) Efectuando cálculos

$$
\begin{aligned}
{}_2E^1_{pq} &= H_p(H_{q-p}(GJ,\partial'),\partial'')\\
&= \frac{Ker\left(H_{q-p}\left(GJ_{*,p};\partial'\right) \xrightarrow{\partial''} H_{q-p}\left(GJ_{*,p-1};\partial'\right)\right)}{Im\left(H_{q-p}\left(GJ_{*,p+1};\partial'\right) \xrightarrow{\partial''} H_{q-p}\left(GJ_{*,p};\partial'\right)\right)}\\
&= \frac{Ker\left({}_2E^0_{pq} \xrightarrow{\partial''} {}_2E^0_{p-1,q-1}\right)}{Im\left({}_2E^0_{p+1,q+1} \xrightarrow{\partial''} {}_2E^0_{pq}\right)}, \text{ por (31) :}\\
&= \frac{Ker\left(G\left(\frac{K_{q-p,p}}{L_{q-p,p}}\right) \xrightarrow{\partial''} G\left(\frac{K_{q-p,p-1}}{L_{q-p,p-1}}\right)\right)}{Im\left(G\left(\frac{K_{q-p,p+1}}{L_{q-p,p+1}}\right) \xrightarrow{\partial''} G\left(\frac{K_{q-p,p}}{L_{q-p,p}}\right)\right)}, \text{ por (25) :}\\
&= L_pG\left(\frac{Z_{q-p}}{B_{q-p}}\right)\\
&= (L_pG)\,(L_{q-p}F)\,(A) \text{ por (23)}, \qquad (32)
\end{aligned}
$$

Se completa los cálculos, hallando $H_q(Tot\ GJ)$, del siguiente modo:
Como $G : \mathfrak{B} \to \mathfrak{C}$ es un funtor covariante aditivo, entonces $L_qG : \mathfrak{B} \to \mathfrak{C}$ es el q-ésimo funtor derivado izquierdo de G y $L_qG\,(FP_r) = H_q\,(GJ_r)$ para $r = 0,1,\ldots$, donde $FP_r \in |\mathfrak{B}|$ y J_r es una resolución proyectiva de FP_r según (26).
Para cada objeto proyectivo P_r de $\mathfrak{A}$, por hipótesis FP_r es G-acíclico, luego se obtiene que

$$L_qG\,(FP_r) = \begin{cases} G(FP_r), & q = 0\\ 0, & q \geq 1 \end{cases} \qquad (33)$$

Así, las homologías de las columnas del complejo doble de cadenas, D, obtenido de (26) aplicando G y haciendo anticonmutativo el diagrama conmutativo, se anulan.



(34)

Por la Proposición 3.5, la homología del complejo total de D se anula; es decir,

$$H_n\,(Tot\,D) = 0,\ \forall n. \qquad (35)$$

Sea D_1 el complejo $Tot\,GJ$ visto como subcomplejo de $Tot\,D$. Entonces se tiene la sucesión exacta corta $GF(\mathcal{P}) \rightarrowtail Tot\ D \twoheadrightarrow D_1$ de complejos de cadenas, donde $(Tot\,D)\,/GF(\mathcal{P}) = D_1$.
El [3, Teorema IV.2.1] asegura que la sucesión larga de homologías siguiente es exacta
$H_0(D_1) \leftarrow H_0(Tot\ D) \leftarrow H_0\,(GF\mathcal{P}) \leftarrow H_1(D_1) \leftarrow H_1(Tot\ D) \leftarrow H_1\,(GF\mathcal{P}) \leftarrow H_2(D_1) \leftarrow H_2(Tot\ D) \leftarrow \cdots$
Por (35) se tiene $H_n(Tot\ D) = 0$ para todo $n = 0,1,\ldots$.
Así, de la sucesión anterior se obtiene $0 \leftarrow H_0\,(GF\mathcal{P}) \xleftarrow{\sim} H_1(D_1) \leftarrow 0 \leftarrow H_1\,(GF\mathcal{P}) \xleftarrow{\sim} H_2(D_1) \leftarrow 0 \leftarrow \cdots$
Como $H_n(D_1) \cong H_{n-1}\,(Tot\ GJ)$ para $n = 1,2,\ldots$, se obtiene sucesivamente:
$L_0\,(GF)\,(A) = H_0\,(GF\mathcal{P}) \cong H_1(D_1) \cong H_0\,(Tot\ GJ)$;
$L_1\,(GF)\,(A) = H_1\,(GF\mathcal{P}) \cong H_2(D_1) \cong H_1\,(Tot\ GJ)$.

$$\text{En general, } H_q\,(Tot\ GJ) = L_q\,(GF)\,(A)\ . \qquad (36)$$

En consecuencia, reemplazando en (28) los valores hallados en (32) y (36), se concluye que para cada objeto A de $\mathfrak{A}$, existe una sucesión espectral $E = {}_2E\,(A)$, satisfaciendo las condiciones requeridas; es decir, tal que $E^1_{pq} = (L_pG)\,(L_{q-p}F)\,(A) \Rightarrow L_q\,(GF)\,(A)$ ■

4 Grupo de Homología con Sucesiones Espectrales.

En esta sección, utilizando la sucesión espectral homológica de Grothendieck se obtiene la versión homológica de Theorem 9.5 de Lyndon-Hochschild-Serre[3], que permite calcular el grupo de homología .

Sea G un grupo escrito multiplicativamente con identidad e. Se denotará
$\mathbb{Z}G = \{r \ / \ r = \sum_{x\in G} m_r(x)x$, para alguna función $m_r : G \to \mathbb{Z}$ con $m_r(x) = 0$ excepto para un número finito de elementos x de $G\}$.

Sea $\imath : G \rightarrowtail \mathbb{Z}G$ la aplicación definida por $\imath(g) = \sum_{x\in G} m_g(x)x$,
donde para cada $g \in G$, $m_g(x) = \begin{cases} 1, & g = x \\ 0, & g \neq x \end{cases}$

La aplicación $\imath : G \rightarrowtail \mathbb{Z}G$ se llama LA INMERSIÓN y claramente $\imath G \subseteq \mathbb{Z}G$, luego $\mathbb{Z}G \neq \emptyset$.
Ahora, considerando $r = \sum_{x\in G} m_r(x)x$ y $s = \sum_{x\in G} m_s(x)x$ se define la suma $r + s$ y el producto rs en $\mathbb{Z}G$, respectivamente, por

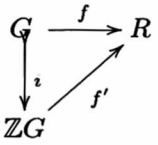
$$r + s = \sum_{x\in G} [m_r(x) + m_s(x)]\, x \tag{37}$$

$$rs = \sum_{x,y\in G} [m_r(x)m_s(y)]\, xy \tag{38}$$

Proposición 4.1. *Dado un grupo G, $\mathbb{Z}G$ con las operaciones dadas en (37) y (38) es un anillo unitario.*

El anillo de la proposición anterior se llama ANILLO DE GRUPO CON COEFICIENTES ENTEROS y se caracteriza por la propiedad universal siguiente.

Proposición 4.2. *Sea G un grupo, $\imath : G \to \mathbb{Z}G$ la inmersión y R un anillo. Para cualquier función $f : G \to R$ con $f(xy) = f(x)f(y)$ y $f(e) = 1_R$ (elemento unitario del anillo R) existe un único morfismo de anillos $f' : \mathbb{Z}G \to R$ tal que $f'\imath = f$.*



Prueba.- Se define $f'\left(\sum_{x\in G} m(x)x\right) = \sum_{x\in G} m(x)f(x)$. Entonces f' es el único morfismo de anillos tal que $f'\imath = f$ ■

Definición 4.3. *Un G-módulo derecho es un grupo abeliano A provisto de un morfismo $\sigma : G \to Aut(A)$ de grupos, definido por $\sigma(x)(a) = ax$, $\forall x \in G$, $\forall a \in A$.*

Cada grupo aditivo abeliano A es un G-módulo derecho (trivial) bajo el morfismo trivial de grupos $\sigma : G \to Aut(A)$ dado por $\sigma(g) = id_A$ para todo $g \in G$.

Proposición 4.4. *Sea G un grupo. Entonces A es un G-módulo derecho si y sólo si A es $\mathbb{Z}G$-módulo derecho.*

Proposición 4.5. *Sea G un grupo y $\mathbb{Z} \otimes_{\mathbb{Z}G} (-) : \mathfrak{m}^l_{\mathbb{Z}G} \to \mathfrak{Ab}$ dado por $A \mapsto \mathbb{Z} \otimes_{\mathbb{Z}G} (A)$. Entonces $\mathbb{Z} \otimes_{\mathbb{Z}G} (-)$ es un funtor covariante aditivo.*

Prueba.- Como el grupo aditivo abeliano $\mathbb{Z}$ es un G-módulo derecho, por Proposición 4.4 $\mathbb{Z}$ es un $\mathbb{Z}G$-módulo derecho. Para $\Lambda = \mathbb{Z}G$, por la [3, Proposición III.7.1] se sabe que $\mathbb{Z} \otimes_{\mathbb{Z}G} (-)$ es un funtor covariante; luego de [3, Proposición II.9.5] y [3, Proposición III.7.3] se deduce que el funtor covariante $\mathbb{Z} \otimes_{\mathbb{Z}G} (-)$ es aditivo ■

Proposición 4.6. *Sea $\varphi : H \to G$ un morfismo de grupos, entonces $\mathbb{Z}\varphi : \mathbb{Z}H \to \mathbb{Z}G$ es un morfismo de anillos definido por $\mathbb{Z}\varphi\left(\sum_{x\in H} m(x)x\right) = \sum_{x\in H} m(x)\varphi(x)$.*

Prueba.- Se verifica que $\mathbb{Z}\varphi$ preserva la suma y el producto ■

Corolario 4.7. *Sea N un subgrupo de un grupo K. Si A es un K-módulo, entonces A es un N-módulo.*

En virtud de Proposición 4.5, dado un grupo G, $\mathbb{Z} \otimes_{\mathbb{Z}G} (-) : \mathfrak{m}^l_{\mathbb{Z}G} \to \mathfrak{Ab}$ es un funtor covariante aditivo, luego para $n \geq 0$ los funtores derivados izquierdos de $\mathbb{Z} \otimes_{\mathbb{Z}G} (-)$ son

$$Tor_n^{\mathbb{Z}G}(\mathbb{Z}, -) = L_n(\mathbb{Z} \otimes_{\mathbb{Z}G} (-)) : \mathfrak{m}^l_{\mathbb{Z}G} \to \mathfrak{Ab}.$$

Por consiguiente está definido $Tor_n^{\mathbb{Z}G}(\mathbb{Z}, A)$ para todo $A \in \left|\mathfrak{m}^l_{\mathbb{Z}G}\right|$.

***Lema* 4.8.** *Sea $N \overset{\imath}{\rightarrowtail} K \overset{p}{\twoheadrightarrow} Q$ una sucesión exacta corta de grupos y A un K-módulo, entonces $\mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ y $\mathbb{Z} \otimes_{\mathbb{Z}K} A$ son $\mathbb{Z}$-módulos izquierdos isomorfos.*

Prueba.- Primero probemos que $\mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ y $\mathbb{Z} \otimes_{\mathbb{Z}K} A$ son $\mathbb{Z}$-módulos izquierdos.
Recordemos que $\mathbb{Z}$ se considera como $\mathbb{Z}G$-módulo trivial izquierdo (derecho) para todo grupo G. En particular, $\mathbb{Z}$ se considera como $\mathbb{Z}K$-módulo trivial derecho.
Puesto que A es un $\mathbb{Z}K$-módulo izquierdo, para el morfismo de anillos unitarios $\varepsilon_K : \mathbb{Z}K \to \mathbb{Z}$ dado por $\varepsilon_K\left(\sum_{x\in K} m(x)x\right) = \sum_{x\in K} m(x)$ (en [3, (VI.1.2)]), por [3, (IV.12.4)] sabemos que $\mathbb{Z} \otimes_{\mathbb{Z}K} A$ es un $\mathbb{Z}$-módulo izquierdo. Similarmente, se prueba que $\mathbb{Z} \otimes_{\mathbb{Z}N} A$ es un $\mathbb{Z}$-módulo izquierdo. Así, $\mathbb{Z} \otimes_{\mathbb{Z}N} A$ es un grupo abeliano.
Puesto que A es un K-módulo, existe un morfismo de anillos $\rho : \mathbb{Z}K \to \mathrm{End}(A)$. Luego
$\mathbb{Z} \otimes_{\mathbb{Z}N} \rho : \mathbb{Z} \otimes_{\mathbb{Z}N} \mathbb{Z}K \to \mathrm{End}(\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ es un morfismo de anillos. Pero $\mathbb{Z} \otimes_{\mathbb{Z}N} \mathbb{Z}K \cong \mathbb{Z}Q$ (ver [8, página 29]), luego $\mathbb{Z} \otimes_{\mathbb{Z}N} A$ es un $\mathbb{Z}Q$-módulo izquierdo. Si consideramos el grupo cociente $Q = K/N$ y el morfismo de anillos unitarios $\varepsilon_Q : \mathbb{Z}Q \to \mathbb{Z}$, resulta que $\mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ es un $\mathbb{Z}$-módulo izquierdo.
Ahora, definamos $\varphi : \mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A) \to \mathbb{Z} \otimes_{\mathbb{Z}N} A$ por

$$\varphi\left(\sum_{i\in I} z_i \otimes (y_i \otimes a_i)\right) = \sum_{i\in I} (z_i y_i) \otimes a_i, \text{ donde } I \text{ es finito;}$$

$\psi : \mathbb{Z} \otimes_{\mathbb{Z}N} A \to \mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ por

$$\psi\left(\sum_{j\in J} z_i \otimes a_i\right) = \sum_{j\in J} z_j \otimes (1 \otimes a_j), \text{ donde } J \text{ es finito.}$$

Sean $b, b_1, b_2 \in \mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)$ y $\lambda \in \mathbb{Z}$, entonces

i) $\varphi(\lambda b) = \lambda\varphi(b)$,

ii) $\varphi(b_1 + b_2) = \varphi(b_1) + \varphi(b_2)$.

En efecto, $\varphi(\lambda b) = \varphi\left(\sum_{i\in I}(\lambda z_i) \otimes (y_i \otimes a_i)\right) = \sum_{i\in I}((\lambda z_i)y_i) \otimes a_i = \lambda\varphi(b)$;

$$\begin{aligned}\varphi(b_1 + b_2) &= \varphi\left(\sum_{i\in I} z_i \otimes (y_i \otimes a_i) + \sum_{j\in J} z_j \otimes (y_j \otimes a_j)\right)\\ &= \varphi\left(\sum_{k\in K} z_k \otimes (y_k \otimes a_k)\right), \ K = I \cup J \text{ finito.}\\ &= \sum_{k\in K}(z_k y_k) \otimes a_k = \sum_{i\in I}(z_i y_i) \otimes a_i\\ &+ \sum_{j\in J}(z_j y_j) \otimes a_j = \varphi(b_1) + \varphi(b_2).\end{aligned}$$

Por lo tanto, φ es un morfismo de $\mathbb{Z}$-módulos.
Similarmente, se prueba que ψ es un morfismo de $\mathbb{Z}$-módulos. Por el hecho que $\psi\varphi = id$ y $\varphi\psi = id$, φ es un isomorfismo de $\mathbb{Z}$-módulos, como se quería probar ■

Definición 4.9. *Sea G un grupo, A un G-módulo. El n-ésimo grupo de homología de G con coeficientes en A denotado por $H_n(G, A)$ se define como*

$$H_n(G, A) = Tor_n^{\mathbb{Z}G}(\mathbb{Z}, A),$$

donde $\mathbb{Z}$ es visto como un G-módulo derecho trivial.
El $\mathbb{Z}$-módulo graduado $H_*(G, A) = \{H_n(G, A)\}$ se llama la homología de G con coeficientes en A.

"Un resultado análogo a [3, Proposición IV.5.3] es la siguiente:"

Proposición 4.10. *Sea $G : \mathfrak{B} \to \mathfrak{C}$ un funtor covariante aditivo entre categorías abelianas, donde $\mathfrak{B}$ tiene suficientes proyectivos y sea $Q \in |\mathfrak{B}|$ proyectivo, entonces $L_nGQ = 0$ para $n = 1, 2, \ldots$ y $L_0GQ \cong GQ$.*

Teorema 4.11 (Lyndon-Hochschild-Serre). *Dada la sucesión exacta corta de grupos*
$N \xrightarrow{\ i\ } K \xrightarrow{\ p\ } Q$ y dado un K-módulo A, existe una sucesión espectral $E = \{E^n(A)\}$ tal que

$$E^1_{pq} = H_p(Q, H_{q-p}(N, A)) \Rightarrow H_q(K, A) \ . \qquad (39)$$

Es decir, la sucesión espectral E converge finitamente a la homología $\{H_q(K, A)\}$, filtrada adecuadamente.

Prueba.- Se realiza aplicando el teorema de la sucesión espectral homológica de Grothendieck, por lo que se verifica las hipótesis correspondientes y esto se hace en los dos items siguientes:

i) Se va a considerar funtores $F : \mathfrak{A} \to \mathfrak{B}$ y $G : \mathfrak{B} \to \mathfrak{C}$ covariantes aditivos entre categorías abelianas, donde $\mathfrak{A}$ y $\mathfrak{B}$ tengan suficientes proyectivos.
Sean $\mathfrak{A}$ la categoría de K-módulos, $\mathfrak{B}$ la categoría de Q-módulos y $\mathfrak{C}$ la categoría de grupos abelianos, entonces $\mathfrak{A} = \mathfrak{m}^l_{\mathbb{Z}K}$, $\mathfrak{B} = \mathfrak{m}^l_{\mathbb{Z}Q}$ y $\mathfrak{C} = \mathfrak{m}^l_{\mathbb{Z}}$. Por [5, p 425] se sigue que las categorías $\mathfrak{A}$, $\mathfrak{B}$ y $\mathfrak{C}$ son abelianas; por otro lado, [3, Proposición I.4.3] indica que las categorías $\mathfrak{A}$ y $\mathfrak{B}$ tienen suficientes proyectivos.
Según Proposición 4.5 para $G = N$ se tiene que $\mathbb{Z} \otimes_{\mathbb{Z}N} (-) : \mathfrak{m}^l_{\mathbb{Z}N} \to \mathfrak{Ab}$ es un funtor covariante aditivo. Pero Corolario 4.7 proporciona la inclusión $\mathfrak{m}^l_{\mathbb{Z}K} \subseteq \mathfrak{m}^l_{\mathbb{Z}N}$, luego definiendo $F = \mathbb{Z} \otimes_{\mathbb{Z}N} (-)|_{\mathfrak{a}} : \mathfrak{A} \to \mathfrak{Ab}$, se nota que para cada $A \in |\mathfrak{A}|$, $F(A) = \mathbb{Z} \otimes_{\mathbb{Z}N} (A)$ es un grupo aditivo abeliano, luego $F(A)$ es un Q-módulo trivial y así $F(A) \in |\mathfrak{B}|$. Por lo tanto $F : \mathfrak{A} \to \mathfrak{B}$ es un funtor covariante aditivo entre categorías abelianas.
Nuevamente, si se define $G = \mathbb{Z} \otimes_{\mathbb{Z}Q} (-) : \mathfrak{B} \to \mathfrak{C}$, aplicando Proposición 4.5 para $G = Q$, se deduce que $G : \mathfrak{B} \to \mathfrak{C}$ es un funtor covariante aditivo entre categorías abelianas.

ii) Si $P \in \left|\mathfrak{m}^l_{\mathbb{Z}K}\right|$ es proyectivo, entonces $F(P) = \mathbb{Z} \otimes_{\mathbb{Z}N} (P)$ es G-acíclico.
Por la Proposición 4.10 debemos demostrar que $F(P) \in \left|\mathfrak{m}^l_{\mathbb{Z}Q}\right|$ es proyectivo siempre que $P \in \left|\mathfrak{m}^l_{\mathbb{Z}K}\right|$ es proyectivo.
Sea $\varphi : \mathbb{Z}N \to \mathbb{Z}$ morfismo de anillos tal que $r = \sum_{x\in N} m_r(x)x \mapsto \sum_{x\in N} m_r(x)$.
Si $U = U_\varphi : \mathfrak{m}^l_{\mathbb{Z}} \to \mathfrak{m}^l_{\mathbb{Z}N}$ es funtor de cambio de anillos, se sabe que $F : \mathfrak{m}^l_{\mathbb{Z}N} \to \mathfrak{m}^l_{\mathbb{Z}}$ dado por $F(A) = \mathbb{Z} \otimes_{\mathbb{Z}N} (A)$ es su adjunto izquierdo. Como U preserva epimorfismos (pág. 319[3]), por Theorem 12.1[3] F preserva proyectivos. Puesto que $P \in \left|\mathfrak{m}^l_{\mathbb{Z}K}\right|$ es proyectivo, se sigue que $F(P) = \mathbb{Z} \otimes_{\mathbb{Z}N} (P)$ es proyectivo. Luego, por la Proposición 4.10 se obtiene

$$L_qG(FP) = \begin{cases} G(FP), & q = 0 \\ 0, & q \geq 1 \end{cases} \qquad (40)$$

Por lo tanto, $F(P)$ es G-acíclico.
De *i)*, *ii)* se ve que se satisfacen las hipótesis del teorema 3.7, entonces para el objeto dado A de $\mathfrak{A}$ existe una sucesión espectral (homológica de Grothendieck) $E = \{E^n(A)\}$ tal que

$$E^1_{pq} = (L_pG)(L_{q-p}F)(A) \Rightarrow L_q(GF)(A) \ . \qquad (41)$$

Es decir, la sucesión espectral E converge finitamente al objeto graduado asociado a $\{L_q(GF)(A)\}$, filtrado adecuadamente. Se completa la prueba, calculando valores de E^1_{pq} y $L_q(GF)(A)$ en los items *iii)* y *iv)*:

iii) Puesto que $F(A) = \mathbb{Z} \otimes_{\mathbb{Z}N} A$ se tiene $(L_{q-p}F)(A) = Tor_{q-p}^{\mathbb{Z}N}(\mathbb{Z}, A) = H_{q-p}(N, A)$, luego

$$\begin{aligned} E_{pq}^1 &= (L_pG)(L_{q-p}F)(A) \\ &= (L_pG)(H_{q-p}(N, A)) \\ &= Tor_p^{\mathbb{Z}Q}(\mathbb{Z}, H_{q-p}(N, A)) = H_p(Q, H_{q-p}(N, A)). \end{aligned}$$

Por lo tanto $E_{pq}^1 = H_p(Q, H_{q-p}(N, A))$. (42)

iv) Del Lema 4.8 obtenemos que $\mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A) \cong \mathbb{Z} \otimes_{\mathbb{Z}K} A$, de modo que

$$\begin{aligned} L_q(GF)(A) &= L_q(\mathbb{Z} \otimes_{\mathbb{Z}Q} (\mathbb{Z} \otimes_{\mathbb{Z}N} A)) \\ &= L_q(\mathbb{Z} \otimes_{\mathbb{Z}K} A) \\ &= Tor_q^{\mathbb{Z}K}(\mathbb{Z}, A) = H_q(K, A). \end{aligned}$$

Luego $L_q(GF)(A) = H_q(K, A)$. (43)

Reemplazando (42) y (43) en (41), se garantiza la existencia de una sucesión espectral $E = \{E^n(A)\}$ tal que $E_{pq}^1 = H_p(Q, H_{q-p}(N, A)) \Rightarrow H_q(K, A)$ ■

Tener una sucesión exacta corta de grupos $N \overset{i}{\rightarrowtail} K \overset{p}{\twoheadrightarrow} Q$ significa tener un subgrupo normal N de K y un grupo cociente $Q = K/N$.

En consecuencia, el grupo de homología $H_*(K, A) = \{H_q(K, A)\}$ de un grupo K con coeficientes en K-módulo A, puede ser aproximado por una sucesión espectral cuyos términos envuelven grupos de homología del grupo cociente Q y del subgrupo normal N.

5 Conclusiones

1) Por definición, el n-ésimo grupo de homología $H_n(G, A) = Tor_n^{\mathbb{Z}G}(\mathbb{Z}, A) = L_n(\mathbb{Z} \otimes_{\mathbb{Z}G})(A)$ de G con coeficientes en A es el valor del satélite izquierdo $Tor^{\mathbb{Z}G}(\mathbb{Z}, -)$ del funtor aditivo $\mathbb{Z} \otimes_{\mathbb{Z}G} (-)$ en un G-módulo A.

2) Por [3, Exercise IV.5.8] un funtor exacto derecho es un funtor aditivo. En consecuencia, la prueba de Teorema 3.7 da una nueva prueba de [5, Corollary 5.8.4].

3) Se puede plantear como materia de nueva investigación que, siguiendo el método aplicado en este artículo, es posible dar una forma de la sucesión espectral de Grothendieck (Teorema 3.7) para funtores $\mathscr{E}$−derivados izquierdos.
Dados $\mathfrak{A}$, $\mathfrak{B}$ categorías abelianas, $\mathscr{E}$ una clase proyectiva de epimorfismos en $\mathfrak{A}$. Sea $T : \mathfrak{A} \to \mathfrak{B}$ un funtor aditivo. Entonces [3, Theorem IX.1.3] garantiza que está bien definido $L_n^{\mathscr{E}}T : \mathfrak{A} \to \mathfrak{B}$ sobre objetos y morfismos, $n = 0, 1, \ldots$, donde $L_n^{\mathscr{E}}T$ es llamado n−ésimo funtor $\mathscr{E}$−derivado izquierdo.
Usando [3, Lemma IX.2.2] y Lema 3.2, se puede establecer lema de herradura para una sucesión $\mathscr{E}$−exacta corta en una categoría $(\mathfrak{A}, \mathscr{E})$.

4) Se encuentra en [9] la extensión de Teorema 3.7 para funtores no aditivos o categorías no abelianas.

Agradecimientos

El presente trabajo es el desarrollo del aspecto homológico de la tesis de maestría del autor y está relacionado con su proyecto de tesis doctoral que está elaborando en IMCA-UNI.
El autor agradece a la Universidad Nacional del Altiplano de Puno por la oportunidad dada para enseñar cursos de álgebra en la Escuela Profesional de Ciencias Físico Matemáticas; al Instituto de Matemática y Ciencias Afines por las condiciones adecuadas que le ofrece para realizar con éxito el doctorado en Matemática.

1. P. Belmans, Spectral Sequences : examples in algebra and algebraic geometry. Lecture Notes in ANAGRAMS Seminar in Spectral Sequences, 2014.
2. D. Popovici, International Journal of Mathematics, **27**(14), 1650111, 2016.
3. P.J. Hilton & U. Stammbach, *A Course in Homological Algebra*, Springer - Verlag New York Heidelberg Berlin, 1971.
4. A. Grothendieck, Tôhoku Math. J. , **9**(2), 119-221, 1957.
5. C.A. Weibel, *An Introduction to Homological Algebra*, Cambridge University Press, 1994.
6. J.J. Rotman, *An Introduction to Homological Algebra*, Springer 1, 2009.
7. G. Tochi, Tesis de Licenciatura, Universidad de Buenos Aires, 2014.
8. C.A. Hurtado, Tesis de Maestría, Pontificia Universidad Católica del Perú, 2016.
9. D. Blanc & C. Stover, "A Generalized Grothendieck Spectral Sequence", in N. Ray and G. Walker, eds., Adams Memorial Symposium on Algebraic Topology, Vol. 1, Lond. Math. Soc. Lec. Notes Ser. 175, Cambridge U. Press Cambridge, 145-161, 1992.

Formulación variacional: Ecuaciones del calor y de onda

Héctor Guimaray Huerta, Eladio Ocaña Anaya
IMCA, Facultad de Ciencias.
Universidad Nacional de Ingeniería;
hguimaray@uni.edu.pe, eocana@imca.uni.edu.pe

Recibido el 6 de Marzo de 2020; aceptado el 25 de Junio de 2020

En este trabajo estudiamos la formulación variacional de las ecuaciones diferenciales parciales del calor y de onda, y posteriormente determinamos, usando los teoremas minimax, la existencia de solución débil de dichas ecuaciones, a través del análisis variacional, considerando estas soluciones como puntos críticos de ciertas funciones definida en un espacio de búsqueda que en nuestro caso será el espacio de Sobolev H^1 ó H_0^1.

Palabras Claves: Fórmula de Green, Soución débil, Punto crítico.

This work deals with the variational formulation of the heat and wave partial differential equations, and then, using the minimax theorems, we study the existence of weak solutions of these equations, through the variational analysis, considering these solutions as critical points of some functions defined on appropriated Sobolev spaces such as H^1 or H_0^1.

Keywords: Green's formula, Weak solution, Critical point.

1 Introducción

Estudiaremos la formulación variacional de las ecuaciones diferenciales parciales del calor y de onda, dando lugar a estudiar la existencia de puntos críticos de ciertas funcionales correspondientes a las ecuaciones diferenciales parciales mencionadas,

Los puntos críticos considerados en el trabajo son los ceros de la derivada de Gateaux [1].

La formulación variacional de una ecuación diferencial parcial (edp) se obtiene a partir de una reformulación integral de tal edp con la intervención de ciertos tipos de funciones llamadas funciones de prueba o test que para nosotros serán las funciones que forman el conjunto $C^\infty(\Omega)$ ó $C_0^\infty(\Omega)$,

$$C_0^\infty(\Omega) := \{f \in C^\infty(\Omega) : \Omega \supset \text{supp}\,(f) \text{ es compacto}\},$$

donde $\text{supp}\,(f) := \overline{\{x \in \Omega : f(x) \neq 0\}}$, siendo Ω abierto de $\mathbb{R}^n$.

Para $m \in \{0\} \cup \mathbb{N}$, definimos el **Espacio de Sobolev** $W^{m,p}(\Omega)$ como [2]:

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) : \exists\, D^\alpha u \in L^p(\Omega),\ 0 \leq |\alpha| \leq m\},$$

donde $D^\alpha u =: g_\alpha$ es la **derivada parcial débil** de u, esto es, la función g_α que satisface

$$\int_\Omega u D^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega g_\alpha \varphi, \quad \forall \varphi \in C_0^\infty(\Omega).$$

Un caso especial de espacio de Sobolev es el espacio $H^m(\Omega)$ y más aún, el espacio $H^1(\Omega)$, definido por:

$$H^m(\Omega) := W^{m,2}(\Omega).$$

Una de las herramientas importantes del análisis variacional es la **Fórmula de Green** [3],

$$\int_\Omega f \triangle g dx = -\int_\Omega \nabla f \nabla g dx + \int_{\partial\Omega} f \frac{\partial g}{\partial \eta} ds,$$

donde $\dfrac{\partial g}{\partial \eta} = \nabla g.\eta$.

Esta identidad nos permite pasar de una edp (búsqueda de solución fuerte) a una ecuación integral (búsqueda de solución débil), y que bajo ciertas condiciones de regualidad, ambas soluciones coinciden.

Nuestro punto de partida en el tranajo es la formulación variacional de la **ecuación de Laplace**:

$$\Delta u = 0, \text{ sobre } \Omega,$$

asociada a las condiciones de frontera, según el caso, de **Dirichlet** o **Neumann**, [4].

2 Ecuaciones de Laplace y Poisson

2.1 Ecuación de Laplace-Dirichlet [5]

La ecuación de Laplace con condición de frontera de Dirichlet viene dada por la siguiente expresión:

$$\begin{cases} \Delta u = 0 & \text{sobre } \Omega, \\ u = 0 & \text{sobre } \partial\Omega. \end{cases} \tag{LD}$$

Una función $u \in C^2(\Omega) \cap C(\overline{\Omega})$ que satisface (LD) es llamada solución fuerte de la ecuación.

2.1.1 Formulación variacional de la ecuación (LD)

De la primera ecuación del sistema (LD), tenemos

$$\varphi \Delta u = 0, \quad \text{para todo } \varphi \in C_0^\infty(\Omega), \tag{1}$$

lo que implica que

$$\begin{aligned} 0 &= \int_\Omega \varphi \triangle u \\ &= -\int_\Omega \nabla\varphi \nabla u + \int_{\partial\Omega} \varphi \frac{\partial u}{\partial \eta} ds \quad \text{(fórmula de Green)} \\ &= -\int_\Omega \nabla u \nabla \varphi. \end{aligned}$$

Así, de la expresión (1) tenemos

$$\int_\Omega \nabla u \nabla \varphi = 0, \quad \text{para todo } \varphi \in C_0^\infty(\Omega). \tag{2}$$

Una función $u \in H_0^1(\Omega)$ que satisface (2) es llamada **solución débil** de la ecuación (LD).

En general, se requieren ciertas condiciones de regularidad sobre $\partial\Omega$ y sobre u para que una solución débil sea también una solución fuerte.

2.1.2 La funcional de energía correspondiente a la ecuación (LD)

Asociada a la ecuación variacional (2), la funcional de energía correspondiente a la ecuación (LD) es la función $\psi : H_0^1(\Omega) \to \mathbb{R}$ definida por

$$\psi(u) = \frac{1}{2}\int_\Omega |\nabla u|^2, \text{ para todo } u \in H_0^1(\Omega). \tag{3}$$

Proposición 2.1. *La función ψ es de clase C^∞ en $H_0^1(\Omega)$. Además,*

$$D_\varphi \psi(u) = \int_\Omega \nabla u \nabla \varphi, \quad \textit{para todo } \varphi \in C_c^\infty(\Omega).$$

Demostración. Siendo $\langle u, v\rangle = \int_\Omega \nabla u \nabla v$ un producto interno de $H_0^1(\Omega)$ cuya norma asociada es $\|u\| = \left(\int_\Omega |\nabla u|^2\right)^{1/2}$, tenemos $\frac{1}{2}\|u\|^2 = \frac{1}{2}\int_\Omega |\nabla u|^2$, de donde

$$\psi(u) = \frac{1}{2}\|u\|^2.$$

Se deduce que ψ es de clase C^∞ en $H_0^1(\Omega)$. Por otra parte,

$$\begin{aligned} D_\varphi \psi(u) &= \lim_{\epsilon\to 0} \frac{\psi(u+\epsilon\varphi) - \psi(u)}{\epsilon} \\ &= \frac{1}{2}\lim_{\epsilon\to 0} \frac{\|u+\epsilon\varphi\|^2 - \|u\|^2}{\epsilon} \\ &= \langle u, \varphi\rangle \\ &= \int_\Omega \nabla u \nabla \varphi, \end{aligned}$$

lo que muestra la identidad deseada. □

Se deduce que toda solución u de (2) es un punto crítico de la función ψ.

2.2 Ecuaciones de Poisson–Dirichlet y Poisson–Neumann [6], [7]

La ecuación de Poisson con condición de frontera de Dirichlet (PD) viene dada por la siguiente expresión:

$$\begin{cases} -\Delta u = f(x,u) & \text{sobre } \Omega, \\ u = g & \text{sobre } \partial\Omega, \end{cases} \tag{PD}$$

para ciertas funciones f y g. Similar al caso anterior, una función $u \in C^2(\Omega) \cap C(\overline{\Omega})$ que satisface (PD) es llamada solución fuerte.

2.2.1 Formulación variacional de la ecuación (PD)

De la primera ecuación del sistema (PD), tenemos

$$-\Delta u \varphi = f(x,u)\varphi, \quad \text{para todo } \varphi \in C_0^\infty(\Omega), \tag{4}$$

lo que implica que

$$\begin{aligned} \int_\Omega f(x,u)\varphi &= \int_\Omega (-\Delta u)\varphi \\ &= -\int_\Omega \varphi \Delta u \\ &= \int_\Omega \nabla\varphi\nabla u - \int_{\partial\Omega} \varphi\frac{\partial u}{\partial\eta} ds \text{ (fórmula de Gre} \\ &= \int_\Omega \nabla u \nabla\varphi. \end{aligned}$$

Por lo tanto,

$$\int_\Omega \nabla u \nabla \varphi = \int_\Omega f(x,u)\varphi, \ \forall \varphi \in C_0^\infty(\Omega). \tag{5}$$

Una función $u \in H_0^1(\Omega)$ que satisface (5) es llamada **solución débil** de la ecuación (PD).

2.2.2 La funcional de energía correspondiente a la ecuación (PD)

Asociada a la ecuación (5), consideremos la función $\Phi : H_0^1(\Omega) \to \mathbb{R}$ definida por

$$\Phi = \psi - \phi, \tag{6}$$

donde ψ es la función definida en (3) y ϕ es la función definida en $H_0^1(\Omega)$ por

$$\phi(u) = \int_\Omega F(x,u)dx, \quad \text{donde} \quad F(x,u) = \int_0^u f(x,s)ds. \tag{7}$$

Tenemos,

$$\begin{aligned} \Phi(u) &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega F(x,u)dx \\ &= \int_\Omega \left[\frac{1}{2}\|\nabla u\|^2 - F(x,u)\right]. \end{aligned}$$

Proposición 2.2. *Se cumple,*

$$D_\varphi \Phi(u) = \int_\Omega \nabla u \nabla \varphi - \int_\Omega f\varphi, \quad \textit{para todo } \varphi \in C_c^\infty(\Omega).$$

Demostración. Es suficiente demostrar que $D_\varphi\phi(u) = \int_\Omega f\varphi$. Tenemos,

$$\begin{aligned} D_\varphi\phi(u) &= \lim_{\epsilon\to 0}\frac{\phi(u+\epsilon\varphi)-\phi(u)}{\epsilon} \\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_\Omega [F(x,u+\epsilon\varphi)-F(x,u)]dx \\ &= \lim_{\epsilon\to 0}\int_\Omega \frac{1}{\epsilon}\left[\int_u^{u+\epsilon\varphi} f(x,s)ds\right]dx \\ &= \int_\Omega \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\int_u^{u+\epsilon\varphi} f(x,s)ds\right]dx \\ &= \int_\Omega \lim_{\epsilon\to 0} f(x,u+\epsilon\varphi)\varphi dx \text{ (regla de L'Hôpital)} \\ &= \int_\Omega f\varphi. \end{aligned}$$

Por lo tanto, $D_\varphi\Phi(u) = \int_\Omega \nabla u\nabla\varphi - \int_\Omega f\varphi$. □

Se deduce que toda solución u de (5) es un punto crítico de Φ.

2.3 La ecuación de Poisson–Neumann

La ecuación de Poisson con condición de frontera de Neumann (PN) viene dada por la siguiente expresión:

$$\begin{cases} -\Delta u = f(x,u), & \text{sobre } \Omega, \\ \dfrac{\partial u}{\partial \eta} = g, & \text{sobre } \partial\Omega, \end{cases} \tag{PN}$$

para ciertas funciones f y g.

2.3.1 Formulación variacional de la ecuación (PN)

De la primera ecuación del sistema (PN), tenemos

$$-\Delta u\varphi = f(x,u)\varphi, \text{ para todo } \varphi \in C^\infty(\overline{\Omega}),$$

de donde

$$\begin{aligned} \int_\Omega f(x,u)\varphi &= -\int_\Omega \varphi\Delta u \\ &= \int_\Omega \nabla\varphi\nabla u - \int_{\partial\Omega}\varphi\frac{\partial u}{\partial\eta}ds \text{ (Green)}. \end{aligned}$$

Una función $u \in H^1(\Omega)$ que satisface la ecuación

$$\int_\Omega \nabla\varphi\nabla u - \int_{\partial\Omega}\varphi g ds = \int_\Omega f(x,u)\varphi,\ \forall\varphi\in C^\infty(\overline{\Omega}), \tag{8}$$

es llamada **solución débil** de la ecuación (PN).

2.3.2 La funcional de energía correspondiente a la ecuación (PN)

Asociada a la ecuación (8), consideremos la función $\Phi : H^1(\Omega) \to \mathbb{R}$ definida por

$$\begin{aligned} \Phi(u) &= -\frac{1}{2}\int_\Omega |\nabla u|^2 + \int_\Omega uf(x,u) - \int_\Omega F(x,u) \\ &= \frac{1}{2}\int_\Omega |\nabla u|^2 - \int_{\partial\Omega} ug - \int_\Omega F(x,u), \end{aligned}$$

donde $F(x,u) = \int_0^u f(x,s)ds$.

Proposición 2.3. *Se cumple,*

$$D_\varphi\Phi(u) = \int_\Omega \nabla\varphi\nabla u - \int_{\partial\Omega} g\varphi ds - \int_\Omega f\varphi,\ \forall\varphi\in C^\infty(\overline{\Omega}).$$

Demostración. Similar a la demostración de la proposición 2.2. □

Se deduce que toda solución u de (8) es un punto crítico de Φ.

3 Ecuación del Calor [8]

Consideremos en este caso el cilindro $Q = \Omega \times (0,\infty)$, donde Ω sigue siendo un conjunto abierto con frontera $\partial\Omega$. La frontera de Q, llamada **frontera lateral de Q**, es el conjunto $\partial Q = \partial\Omega \times (0,\infty)$.

La ecuación del calor con condición de frontera de Dirichlet viene dada por la siguiente expresión:

$$\begin{cases} u_t - \Delta u = f(x,t,u) & \text{sobre } Q, \\ u = g(x.t,u) & \text{sobre } \partial Q, \\ u(x,0) = u_0(x) & \text{sobre } \Omega, \end{cases} \tag{CD}$$

para ciertas funciones f, g y u_0.

3.0.1 Formulación variacional de la ecuación (CD)

De la primera ecuación de (CD) tenemos

$$u_t\varphi - \Delta u\varphi = f\varphi, \text{ para todo } \varphi \in C^\infty(\overline{Q}),$$

de donde

$$\begin{aligned} \int_Q f\varphi &= \int_Q u_t\varphi - \int_Q \varphi\Delta u \\ &= \int_Q u_t\varphi + \int_Q \nabla u\nabla\varphi - \int_{\partial Q}\frac{\partial u}{\partial\eta}\varphi \text{ (Green)} \\ &= \int_Q u_t\varphi + \int_Q \nabla u\nabla\varphi - \int_{\partial Q}\frac{\partial g}{\partial\eta}\varphi. \end{aligned}$$

Por lo tanto,

$$\int_Q u_t\varphi + \int_Q \nabla u\nabla\varphi = \int_Q f\varphi + \int_{\partial Q}\frac{\partial g}{\partial\eta}\varphi,\ \forall\varphi\in C^\infty(\overline{Q}). \tag{9}$$

Una función u en $H^1(Q)$ que satisface (9) es llamada solución débil de la ecuación (CD).

3.0.2 La funcional de energía correspondiente a la ecuación (CD)

Asociada a la ecuación (9) consideremos la función $\Phi : H^1(Q) \to \mathbb{R}$ definida por

$$\Phi = \psi - \phi + \Psi - \Theta, \tag{10}$$

donde ψ y ϕ son las funciones definidas en (3) y (7), respectivamente, y Ψ y Θ son las funciones definidas en $H^1(Q)$ por:

$$\begin{aligned} \Psi(u) &= \int_Q \mathcal{F}(x,t,u)dx, \text{ donde} \\ \mathcal{F}(x,t,u) &= \int_0^u u_t(x,t,s)ds, \end{aligned} \tag{11}$$

y

$$\Theta(u) = \int_{\partial Q} \mathcal{G}(x,t,u)dx, \text{ con } \mathcal{G}(x,t,u) = \int_0^u \frac{\partial g}{\partial \eta}(x,t,s)ds. \tag{12}$$

Tenemos,

$$\Phi(u) = \int_Q \left[\frac{1}{2}\|\nabla u\|^2 - F(x,t,u) + \mathcal{F}(x,t,u)\right] - \int_{\partial Q} \mathcal{G}(x,t,u).$$

Proposición 3.1. *Se cumple,*

$$D_\varphi \Phi(u) = \int_Q \nabla u \nabla \varphi - \int_Q f\varphi + \int_Q u_t\varphi - \int_{\partial Q} \frac{\partial g}{\partial \eta}\varphi,$$

para todo $\varphi \in H^1(Q)$.

Demostración. enemos

$$\begin{aligned} D_\varphi \Psi(u) &= \lim_{\epsilon\to 0} \frac{\Psi(u+\epsilon\varphi) - \Psi(u)}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_Q [\mathcal{F}(x,t,u+\epsilon\varphi) - \mathcal{F}(x,t,u)]dx \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_Q \left[\int_u^{u+\epsilon\varphi} u_t(x,t,s)ds\right] dx \\ &= \int_Q \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\int_u^{u+\epsilon\varphi} u_t(x,t,s)ds\right] dx \\ &= \int_Q \lim_{\epsilon\to 0} u_t(x,t,u+\epsilon\varphi)\varphi dx, \text{ (L'Hôpital)} \\ &= \int_Q u_t\varphi. \end{aligned}$$

Asimismo,

$$\begin{aligned} D_\varphi \Theta(u) &= \lim_{\epsilon\to 0} \frac{\Theta(u+\epsilon\varphi) - \Theta(u)}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\partial Q} [\mathcal{G}(x,t,u+\epsilon\varphi) - \mathcal{G}(x,t,u)]dx \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\partial Q} \left[\int_u^{u+\epsilon\varphi} \frac{\partial g}{\partial \eta}(x,t,s)ds\right] dx \\ &= \int_{\partial Q} \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\int_u^{u+\epsilon\varphi} \frac{\partial g}{\partial \eta}(x,t,s)ds\right] dx \\ &= \int_{\partial Q} \lim_{\epsilon\to 0} \frac{\partial g}{\partial \eta}(x,t,u+\epsilon\varphi)\varphi dx, \text{ (L'Hôpital)} \\ &= \int_{\partial Q} \frac{\partial g}{\partial \eta}\varphi. \end{aligned}$$

Por lo tanto, $D_\varphi\Phi(u) = \int_Q \nabla u\nabla\varphi - \int_Q f\varphi + \int_Q u_t\varphi - \int_{\partial Q} \frac{\partial g}{\partial \eta}\varphi$. □

Se deduce que toda solución u de (9) es un punto crítico de Φ.

4 Ecuación de Onda [9]

Similar a la ecuación del calor, consideremos $Q = \Omega \times (0,\infty)$, siendo Ω abierto con frontera $\partial\Omega$ y $\partial Q = \partial\Omega \times (0,\infty)$.

La ecuación de onda con condición de frontera de Dirichlet viene dada por la siguiente expresión:

$$\begin{cases} u_{tt} - \Delta u = f(x,t,u) & \text{sobre } Q, \\ u = g(x,t,u) & \text{sobre } \partial Q, \\ u(x,0) = u_0(x) & \text{sobre } \Omega, \\ \frac{\partial u}{\partial t}(x,0) = \nu_0(x) & \text{sobre } \Omega, \end{cases} \tag{OD}$$

para ciertas funciones f, g, u_0 y ν_0.

4.0.1 Formulación variacional de la ecuación (OD)

De la primera ecuación de (OD) tenemos

$$u_{tt}\varphi - \Delta u\varphi = f\varphi, \text{ para todo } \varphi \in H^1(Q),$$

de donde

$$\begin{aligned} \int_Q f\varphi &= \int_Q (u_{tt} - \Delta u)\varphi \\ &= \int_Q u_{tt}\varphi - \int_Q \varphi\Delta u \\ &= \int_Q u_{tt}\varphi + \int_Q \nabla u\nabla\varphi - \int_{\partial Q} \frac{\partial u}{\partial \eta}\varphi ds. \end{aligned}$$

Por lo tanto,

$$\int_Q u_{tt}\varphi + \int_Q \nabla u\nabla\varphi = \int_Q f\varphi + \int_{\partial Q} \frac{\partial g}{\partial \eta}\varphi, \ \forall \varphi \in H^1(\Omega) \tag{13}$$

Una función u en $H^1(Q)$ que satisface (13) es llamada solución débil de la ecuación (OD).

4.0.2 La funcional de energía correspondiente a la ecuación (OD)

Asociada a la ecuación (13), consideremos la función $\Phi : H^1(Q) \to \mathbb{R}$ definida por

$$\Phi = \psi - \phi + \Xi - \Theta, \tag{14}$$

donde ψ, ϕ y Θ son aquellas definidas en (3), (7) y (12). respectivamete, y Ξ es la función definida en $H^1(Q)$ por

$$\Xi(u) = \int_Q \mathcal{F}(t,x,u), \text{ donde } \mathcal{F}(t,x,u) = \int_0^u u_{tt}(t,x,s)ds. \tag{15}$$

Tenemos,

$$\begin{aligned}\Phi(u) &= \frac{1}{2}\int_Q \|\nabla u\|^2 - \int_Q F(t,x,u) + \int_Q \mathcal{F}(t,x,u) - \\ &\quad \int_{\partial Q} \mathcal{G}(x,t,u) \\ &= \int_Q \left[\frac{1}{2}\|\nabla u\|^2 - F(t,x,u) + \mathcal{F}(t,x,u)\right] - \\ &\quad \int_{\partial Q} \mathcal{G}(x,t,u).\end{aligned}$$

Proposición 4.1. *Se cumple,*

$$D_\varphi \Phi(u) = \int_\Omega \nabla u \nabla \varphi - \int_\Omega f\varphi + \int_\Omega u_{tt}\varphi - \int_{\partial Q} \frac{\partial g}{\partial \eta}\varphi.$$

Demostración. Tenemos

$$\begin{aligned}D_\varphi \Xi(u) &= \lim_{\epsilon\to 0} \frac{\Xi(u+\epsilon\varphi) - \Xi(u)}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_Q [\mathcal{F}(t,x,u+\epsilon\varphi) - \mathcal{F}(t,x,u)] \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_Q \left[\int_u^{u+\epsilon\varphi} u_{tt}(t,x,s)ds\right] \\ &= \int_Q \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\int_u^{u+\epsilon\varphi} u_{tt}(t,x,s)ds\right] \\ &= \int_Q \lim_{\epsilon\to 0} u_{tt}(t,x,u+\epsilon\varphi)\varphi, \text{ (regla de L'Hôpital)} \\ &= \int_Q u_{tt}\varphi.\end{aligned}$$

Por lo tanto, $D_\varphi\Phi(u) = \int_Q \nabla u\nabla\varphi - \int_Q f\varphi + \int_Q u_{tt}\varphi - \int_{\partial Q}\frac{\partial g}{\partial \eta}\varphi$. □

Se deduce que toda solución u de (13) es un punto crítico de Φ.

5 Conclusiones

A través del análisis variacional hemos estudiado las formulaciones variacionales de las edps del calor y de onda. Las soluciones de estas formulaciones variacionales coinciden con los ceros de la derivada de Gateaux de ciertas funciones, llamadas funciones de energía.

1. Troyanski, S. L., *Gateaux differentiable norms in Lp*, Math. Ann. 287, 221-227, 1990.
2. Krantz, Steven G., *Partial Differential Equations and Complex Analysis*, CRC Press, Inc., USA, 1992.
3. Evans, Laurence C., *Partial Differential Equations*, University of California, USA, 1998.
4. Epstein, Marcelo, *Partial Differential Equations. Mathematical Techniques for Engineers*, Springer International Publishing, Switzerland, 2017.
5. Kartashov, E. M., 57, 13, 1149-1155, 2010.
6. Chang, Jen-Shih, *Handbook of Electrostatic Processes*, Marcel Dekker, Inc., New York, 1995
7. Molchanov, I. N., Zh. Vychisl. Mat. Mat. Fiz., 13, 6, 1607-1612, 1973.
8. Brezis, Haim, *Functional Analysis, Sovolev Spaces and Partial Differential Equations*, Springer, New York, 2011.
9. Keller, J. B., *On solutions of nonlinear wave equations*, Communications on Pure and Applied Mathematics, Vol. X, 523-530, 1957.

Validez de la formulación variacional y existencia de solución débil

Héctor Guimaray Huerta, Eladio Ocaña Anaya
IMCA, Facultad de Ciencias.
Universidad Nacional de Ingeniería;
hguimaray@uni.edu.pe, eocana@imca.uni.edu.pe

Recibido el 27 de Marzo de 2020; aceptado el 17 de Junio de 2020

Estudiaremos algunas condiciones para garantizar la validez tanto de las formulaciones variacionales como de las funcionales asociadas a algunas ecuaciones diferenciales parciales. Asimismo, estudiaremos la existencia de solución de algunas ecuaciones particulares formuladas.

Palabras Claves: Formulación variacional, Soución débil.

We will study some conditions ensuring the validity of the variational formulations as well as the corresponding functional associated to some partial differential equations. We will also study the existence of solution of some particular variational formulations.

Keywords: Variational formulation, Weak solution.

1 Introducción

Sean X e Y dos espacios vectoriales normados, denotamos por $\mathcal{L}(X,Y)$ al espacio lineal

$$\mathcal{L}(X,Y) = \{\Phi : \Phi : X \to Y \text{ es lineal}\}.$$

Se dice que $f : A \subseteq X \to Y$ es **Fréchet diferenciable** [1] en $a \in A$ si existe $L \in \mathcal{L}(X,Y)$ continua tal que

$$\lim_{x\to 0,\, x\in\mathcal{F}} \frac{f(a+x) - f(a) - L(x)}{\|x\|} = 0,$$

donde $\mathcal{F} = \{x \in X : a + x \in A\}$. L es llamada la derivada de Fréchet de f en a y será denotada por $Df(a)$.

La función f es **Gateaux diferenciable** [2] en $a \in A$ si existe $L \in \mathcal{L}(X,Y)$ continua tal que

$$\lim_{t\to 0} \frac{f(a+tv) - f(a) - tL(v)}{t} = 0,\ \forall v \in X \text{ con } a+tv \in A.$$

L es llamada la **derivada de Gateaux** de f en a y será denotada por $f'(a)$.

La derivada direccional de f en el punto a y en la dirección v, es el límite (si existe)

$$D_v f(a) = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t}.$$

En general se cumple $D_v f(a) = f'(a)v$.

La formulación variacional de una ecuación diferencial parcial se obtiene a partir de una reformulación integral de tal ecuación con el uso de ciertos tipos de funciones llamadas funciones test que serán las funciones que forman el conjunto $C^\infty(\Omega)$ ó $C_0^\infty(\Omega)$ [3],

$$C_0^\infty(\Omega) := \{f \in C^\infty(\Omega) : \Omega \supset \operatorname{supp}(f) \text{ es compacto}\},$$

donde $\operatorname{supp}(f) := \overline{\{x \in \Omega : f(x) \neq 0\}}$, siendo Ω abierto de $\mathbb{R}^n$.

Para $m \in \{0\} \cup \mathbb{N}$, definimos el **Espacio de Sobolev** $W^{m,p}(\Omega)$ [4] como:

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) : \exists\, D^\alpha u \in L^p(\Omega),\ 0 \le |\alpha| \le m\},$$

donde $D^\alpha u =: g_\alpha$ es la **derivada parcial débil** de u, esto es, la función g_α que satisface

$$\int_\Omega u D^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega g_\alpha \varphi, \quad \forall \varphi \in C_0^\infty(\Omega).$$

Un caso especial de espacio de Sobolev es el espacio $H^m(\Omega)$ y más aún, el espacio $H^1(\Omega)$, definido por:

$$H^m(\Omega) := W^{m,2}(\Omega).$$

Teorema 1.1 (Fórmula de Green [5]). *Sean $f, g \in C^2(\overline{\Omega})$, entonces*

$$\int_\Omega f \triangle g dx = -\int_\Omega \nabla f \nabla g dx + \int_{\partial\Omega} f \frac{\partial g}{\partial \eta} ds,$$

donde $\frac{\partial g}{\partial \eta} = \nabla g.\eta$

Definición 1.1 (Condición Palais–Smale [6],[1]). *Sean X un espacio de Banach y $\Phi : X \to \mathbb{R}$ de clase C^1. Se dice que Φ satisface la condición* **Palais–Smale (PS)** *si toda sucesión $\{x_k\}$ en X satisfaciendo las condiciones:*

$$\begin{cases} \{\Phi(x_k)\} \text{ acotada, y} \\ \Phi'(x_k) \to 0, \end{cases} \tag{1}$$

admite una subsucesión convergente.

1.1 La ecuación de Laplace–Dirichlet [7]

La ecuación de Laplace con condición de frontera de Dirichlet viene dada por la siguiente expresión:

$$\begin{cases} \Delta u = 0 & \text{sobre } \Omega, \\ u = 0 & \text{sobre } \partial\Omega. \end{cases} \tag{LD}$$

Una función $u \in C^2(\Omega) \cap C(\overline{\Omega})$ que satisface (LD) es llamada solución fuerte de la ecuación.

1.1.1 Formulación variacional de la ecuación (LD)

De la primera ecuación del sistema (LD), tenemos

$$\varphi \Delta u = 0, \quad \text{para todo } \varphi \in C_0^\infty(\Omega), \tag{1}$$

lo que implica que

$$\begin{aligned} 0 &= \int_\Omega \varphi \triangle u \\ &= -\int_\Omega \nabla\varphi \nabla u + \int_{\partial\Omega} \varphi \frac{\partial u}{\partial \eta} ds \quad \text{(fórmula de Green)} \\ &= -\int_\Omega \nabla u \nabla \varphi. \end{aligned}$$

Así, de la expresión (1) tenemos

$$\int_\Omega \nabla u \nabla \varphi = 0, \quad \text{para todo } \varphi \in C_0^\infty(\Omega). \tag{2}$$

Una función $u \in H_0^1(\Omega)$ que satisface (2) es llamada **solución débil** de la ecuación (LD).

1.1.2 La funcional de energía correspondiente a la ecuación (LD)

Asociada a la ecuación variacional (2), la funcional de energía correspondiente a la ecuación (LD) es la función $\psi : H_0^1(\Omega) \to \mathbb{R}$ definida por

$$\psi(u) = \frac{1}{2} \int_\Omega |\nabla u|^2, \text{ para todo } u \in H_0^1(\Omega). \tag{3}$$

1.2 La ecuación de Poisson–Dirichlet [8]

La ecuación de Poisson con condición de frontera de Dirichlet (PD) viene dada por la siguiente expresión:

$$\begin{cases} -\Delta u = f(x,u) & \text{sobre } \Omega, \\ u = g & \text{sobre } \partial\Omega, \end{cases} \tag{PD}$$

para ciertas funciones f y g. Similar al caso anterior, una función $u \in C^2(\Omega) \cap C(\overline{\Omega})$ que satisface (PD) es llamada solución fuerte.

1.2.1 Formulación variacional de la ecuación (PD)

De la primera ecuación del sistema (PD), tenemos

$$-\Delta u \varphi = f(x,u)\varphi, \quad \text{para todo } \varphi \in C_0^\infty(\Omega), \tag{4}$$

lo que implica que

$$\begin{aligned} \int_\Omega f(x,u)\varphi &= \int_\Omega (-\Delta u)\varphi \\ &= -\int_\Omega \varphi \Delta u \\ &= \int_\Omega \nabla\varphi \nabla u - \int_{\partial\Omega} \varphi \frac{\partial u}{\partial \eta} ds \text{ (Green)} \\ &= \int_\Omega \nabla u \nabla \varphi. \end{aligned}$$

Por lo tanto,

$$\int_\Omega \nabla u \nabla \varphi = \int_\Omega f(x,u)\varphi, \ \forall \varphi \in C_0^\infty(\Omega). \tag{5}$$

Una función $u \in H_0^1(\Omega)$ que satisface (5) es llamada **solución débil** de la ecuación (PD).

1.2.2 La funcional de energía correspondiente a la ecuación (PD)

Asociada a la ecuación (5), consideremos la función $\Phi : H_0^1(\Omega) \to \mathbb{R}$ definida por

$$\Phi = \psi - \phi, \tag{6}$$

donde ψ es la función definida en (3) y ϕ es la función definida en $H_0^1(\Omega)$ por

$$\phi(u) = \int_\Omega F(x,u)dx, \text{ donde } F(x,u) = \int_0^u f(x,s)ds. \tag{7}$$

Tenemos,

$$\begin{aligned} \Phi(u) &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega F(x,u)dx \\ &= \int_\Omega \left[\frac{1}{2}\|\nabla u\|^2 - F(x,u)\right]. \end{aligned}$$

Teorema 1.2 ([9]). *Sean X un espacio de Banach, $\Phi : X \to \mathbb{R}$ de clase C^1 e inferiormente acotada. Si Φ satisface la condición PS, entonces esta alcanza su mínimo global.*

2 Validez de formulaciones variacionales y funcionales correspondientes

Proposición 2.1. *Sea $\Omega \subset \mathbb{R}^n$ $(n \geq 3)$ abierto y acotado (con condición de regularidad sobre el borde $\partial\Omega$). Sea $f : \mathbb{R} \to \mathbb{R}$ continua satisfaciendo la siguiente condición: Existen $a, b > 0$ tales que*

$$|f(t)| \leq a + b|t|^{2^*-1} \quad \text{para todo } t \in \mathbb{R}, \tag{8}$$

donde $2^* := \dfrac{2n}{n-2}$ (conocido como el exponente crítico de Sobolev). Sean $F : \mathbb{R} \to \mathbb{R}$ definida por $F(t) = \displaystyle\int_0^t f(s)ds$ y $J : H^1(\Omega) \to \mathbb{R}$ definida por

$$J(u) = \int_\Omega F(u(x))dx.$$

Entonces J es Fréchet diferenciable en $H^1(\Omega)$ y

$$J'(u)v = \int_\Omega f(u(x))v(x)dx \quad \textit{para todo } u, v \in H^1(\Omega).$$

Lo mismo sucede si la funcional J está definida en $H_0^1(\Omega)$ (sin ninguna condición de regularidad sobre $\partial\Omega$).

Demostración. Notemos en primer lugar que la condición (8) implica que J está bien definida.

Para demostrar que J es Fréchet diferenciable, primero demostraremos que J' es Gateaux diferenciable y luego, su derivada de Gateaux J', es continua.

Recordemos la siguiente identidad elemental: para todo $p > 0$, existe $c_p > 0$ tal que

$$|a+b|^p \le c_p(|a|^p + |b|^p) \text{ para todo } a, b \in \mathbb{R}.$$

i) Demostraremos que para todo $u, v \in H^1(\Omega)$,

$$\lim_{t\to 0}\int_\Omega \frac{F(u+tv)-F(u)}{t}dx = \lim_{t\to 0}\int_\Omega f(u)vdx.$$

Cláramente, para todo $x \in \Omega$,

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

Por el teorema de valor medio, existe $\theta \in \mathbb{R}$ con $|\theta| \le |t|$, tal que

$$\begin{aligned} |q(t)| &= [f(u(x)+\theta v(x))v(x)| \\ &\le \left(a + b|u(x)+\theta v(x)|^{2^*-1}\right)|v(x)| \\ &\le c_1|v(x)| + c_2|u(x)|^{2^*-1}|v(x)| + c_3|v(x)|^{2^*} \end{aligned}$$

donde

$$q(t) = \frac{F(u(x)+tv(x))-F(u(x))}{t},$$

para ciertas constantes positivas c_1, c_2 y c_3. El lado derecho de la última desigualdad está en $L^1(\Omega)$ y por lo tanto, por el teorema de la convergencia dominada, tenemos

$$\lim_{t\to 0}\int_\Omega \frac{F(u+tv)-F(u)}{t}dx = \lim_{t\to 0}\int_\Omega f(u)vdx.$$

Como el lado derecho, como función de v, es lineal y continua en $H^1(\Omega)$, J es Gateaux diferenciable.

ii) Demostremos que $J' : H^1(\Omega) \to [H^1(\Omega)]'$ es continua. Sea $\{u_k\}$ en $H^1(\Omega)$ tal que $u_k \to u$ en $H^1(\Omega)$. Siendo $H^1(\Omega)$ inmerso en $L^{2^*}(\Omega)$, existe una subsucesión (que lo denotaremos del mismo modo) $\{u_k\}$ tal que

- $u_k \to u$ en $L^{2^*}(\Omega)$;
- $u_k(x) \to u(x)$ en ctp de Ω;
- existe $w \in L^{2^*}(\Omega)$ tal que $u_k(x) \le w(x)$ en ctp de Ω y para todo k.

Por la desigualdad de Hölder, tenemos

$$|(J'(u_k)-J'(u))v| \le \int_\Omega |f(u_k)-f(u)||v|dx$$

$$\le \left(\int_\Omega |f(u_k)-f(u)|^{\frac{2^*}{2^*-1}}\right)^{\frac{2^*-1}{2^*}}\left(\int_\Omega |v|^{2^*}\right)^{\frac{1}{2^*}}$$

De otro lado, como $\lim_{k\to\infty}|f(u_k(x))-f(u(x))| = 0$ en ctp de Ω, y

$$\begin{aligned} |f(u_k)-f(u)|^{\frac{2^*}{2^*-1}} &\le c\left(1+|u_k|^{2^*-1}+|u|^{2^*-1}\right)^{\frac{2^*}{2^*-1}} \\ &\le C\left(1+|w|^{2^*-1}+|u|^{2^*-1}\right)^{\frac{2^*}{2^*-1}} \\ &\le C\left(1+|w|^{2^*}+|u|^{2^*}\right) \in L^1(\Omega). \end{aligned}$$

Por el teorema de la convergencia dominada,

$$\lim_{k\to\infty}\int_\Omega |f(u_k)-f(u)|^{\frac{2^*}{2^*-1}} = 0$$

y por lo tanto,

$$\begin{aligned} |(J'(u_k)-J'(u))v| &= \sup[\,|(J'(u_k)-J'(u))v| : v \in S_1\,] \\ &\le C\left(\int_\Omega |f(u_k)-f(u)|^{\frac{2^*}{2^*-1}}\right)^{\frac{2^*-1}{2^*}} \end{aligned}$$

donde $S_1 = \{v \in H^1(\Omega), \|v\| = 1\}$. Se deduce que $J'(u_k) \to J'(u)$ en $[H^1(\Omega)]'$, lo que implica que J es Fréchet diferenciable en $H^1(\Omega)$ y $J'(u)v = \int_\Omega f(u)vdx$. □

Definición 2.1 (Función Carathéodory [10]). *Sea $\Omega \subset \mathbb{R}^n$ abierto y acotado. Se dice que una función $f : \Omega \times \mathbb{R} \to \mathbb{R}$ es Carathéodory si satisface las siguientes propiedades:*

i) $\forall s \in \mathbb{R}$, $f(\cdot, s)$ es medible en Ω;

ii) $\forall x \in \Omega$, $f(x, \cdot)$ es continua en $\mathbb{R}$.

Similar a la proposición 2.1, consideremos la siguiente condición de crecimiento: existen $c, d > 0$ y $1 \le \alpha \le 2^*$ (si $n \ge 3$) ó $1 \le \alpha < \infty$ (si $n = 1, 2$) tales que

$$|f(x,s)| \le c + d|s|^\alpha, \text{ para todo } t \in \mathbb{R}. \qquad (9)$$

Proposición 2.2 ([10]). *Sea $f : \Omega\times\mathbb{R} \to \mathbb{R}$ una función Carathéodory satisfaciendo la condición de crecimiento (9) y $F : \Omega \times \mathbb{R} \to \mathbb{R}$ definida por*

$$F(x,s) = \int_0^s f(x,\tau)d\tau.$$

Entonces la funcional $\Phi : H_0^1(\Omega) \to \mathbb{R}$ definida por

$$\Phi(u) = \int_\Omega F(x,u)dx,$$

es de clase C^1 en H_0^1, con

$$D\Phi(u)v = \int_\Omega f(x,u)v \quad \textit{para todo } u, v \in H_0^1(\Omega).$$

3 Existencia de solución débil

En esta sección estudiaremos la existencia de solución de algunas ecuaciones particulares que se relacionan a las ecuaciones formuladas en la Introducción.

Comencemos considerando la ecuación de Poisson–Dirichlet (PD) donde la función f depende solamente de $x \in \Omega$.

Proposición 3.1. *Sea $\Omega \subseteq \mathbb{R}^n$ abierto acotado y $f \in L^2(\Omega)$. La función Φ definida en (6) correspondiente a la ecuación (PD) está bien definida y satisface la condición Palais–Smale.*

Demostración. Por definición, $F(x,u) = \int_0^u f(x)ds = f(x)u$, de donde

$$\begin{aligned}\Phi(u) &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega F(x,u)dx \\ &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega f(x)udx,\end{aligned}$$

Tenemos

$$\begin{aligned}\langle \Phi'(u), \varphi\rangle &= \Phi'(u)\varphi \\ &= D_\varphi \Phi(u) \\ &= \int_\Omega \nabla u\nabla\varphi - \int_\Omega f(x)\varphi \quad \forall \varphi \in H_0^1(\Omega).\end{aligned}$$

Siendo $\langle \Phi'(u), \varphi\rangle = \int_\Omega \nabla(\Phi'(u))\nabla\varphi$ (producto interno en $H_0^1(\Omega)$), tenemos

$$\int_\Omega \nabla(\Phi')\nabla\varphi = \int_\Omega \nabla u\nabla\varphi - \int_\Omega f(x)\varphi \; \forall \varphi \in H_0^1(\Omega). \quad (10)$$

Por la fórmula de Green,

$$\begin{aligned}\int_\Omega \nabla(\Phi'(u))\nabla\varphi &= \int_\Omega \nabla(\Phi'(u))\nabla\varphi - \int_{\partial\Omega} \varphi\frac{\partial\Phi'(u)}{\partial\eta}ds \\ &= -\int_\Omega \varphi\Delta(\Phi'(u))\end{aligned}$$

de donde

$$\int_\Omega \nabla(\Phi'(u))\nabla\varphi = -\int_\Omega \varphi\Delta(\Phi'(u)).$$

Análogamente,

$$\int_\Omega \nabla u\nabla\varphi = -\int_\Omega \varphi\Delta u.$$

Así, de (10), tenemos

$$\begin{aligned}-\int_\Omega \varphi\Delta(\Phi'(u)) &= -\int_\Omega \varphi\Delta u - \int_\Omega f(x)\varphi \\ &= \int_\Omega (-\Delta u - f(x))\varphi \quad \forall\varphi \in H_0^1(\Omega),\end{aligned}$$

de donde

$$-\Delta(\Phi'(u)) = -\Delta u - f(x),$$

lo que equivale a

$$u = \Phi'(u) + (-\Delta)^{-1}f(x).$$

Si $\{u_k\}$ es una sucesión en $H_0^1(\Omega)$ satisfaciendo $\Phi(u_k)$ acotada y $\Phi'(u_k) \to 0$, tenemos,

$$u_k = \Phi'(u_k) + (-\Delta)^{-1}f(x),$$

lo que implica (aún sin asumir la condición $\Phi(u_k)$ acotada) que

$$u_k \to (-\Delta)^{-1}f(x).$$

Se deduce que una subsucesión de $\{u_k\}$ (de hecho toda la sucesión) es convergente. Por lo tanto, Φ satisface la condición Palais-Smale. □

Proposición 3.2. *Si además de las condiciones de la proposición anterior, la función g es cero, entonces la función correspondiente Φ alcanza su mínimo global en $H_0^1(\Omega)$. En particular, la función ψ definida en (3) correspondiente a la ecuación de Laplace–Dirichlet (LD) alcanza su mínimo global en $H_0^1(\Omega)$.*

Demostración. Teniendo en cuenta la proposición 3.1 sobre la propiedad PS de la función Φ y el teorema 1.2 sobre la existencia de mínimo global, mostraremos que la función Φ es acotada inferiormente. Siendo Φ de la forma (10):

$$\Phi(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega f(x)udx,$$

tenemos

$$\Phi(u) \geq -\int_\Omega f(x)udx. \quad (11)$$

Por la desigualdad de Poincaré, tenemos $\|u\|_2 \leq c\|\nabla u\|$ (siendo c una constante positiva), lo que implica

$$\begin{aligned}\|u\|_2^2 &\leq c^2\|\nabla u\|_2^2 \\ &= c^2\langle \nabla u, \nabla u\rangle \\ &= c^2\int_\Omega \nabla u\nabla u \\ &= c^2(-\int_\Omega u\Delta u + \int_{\partial\Omega} u\frac{\partial u}{\partial\eta}), \text{ (fórmula de Green)} \\ &= c^2\int_\Omega (-\Delta u)u \\ &= c^2\int_\Omega fu \\ &\leq c^2\|f\|\|u\|.\end{aligned}$$

Se deduce que $\|u\| \leq c^2\|f\|$ y por lo tanto, por la desigualdad (11), tenemos

$$\Phi(u) \geq -c^2\|f\|^2,$$

de donde Φ es acotada inferiormente. Por el Teorema 1.2, la función Φ alcanza su mínimo global en $H_0^1(\Omega)$. □

Ahora asumiremos que la función f de la ecuación de Poisson–Dirichlet (PD), depende solamente de u:

$$\begin{cases}-\Delta u = f(u), & x \in \Omega \\ u = 0, & x \in \partial\Omega\end{cases} \quad (12)$$

Proposición 3.3 ([11]). *Sean $\Omega \subset \mathbb{R}^n$ abierto y acotado y $f : \mathbb{R} \to \mathbb{R}$ continua satisfaciendo las siguientes propiedades:*

- $|f(u)| \leq a + b\|u\|^{\frac{n+2}{n-2}}$, *con* $a, b \in \mathbb{R}$ *y* $n \geq 3$;
- $f(t)t \leq 0$; *y*
- $[f(t) - f(s)](t - s) \leq 0$ *para todo* $t, s \in \mathbb{R}$.

Entonces la funcional Φ *definida en (6) correspondiente a la ecuación (12) admite un único mínimo en* $H_0^1(\Omega)$.

Demostración. La funcional Φ tiene la siguiente expresión

$$\Phi(u) = \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega F(u)dx,$$

donde $F(u) = \int_0^u f(s)ds$. Se observa que Φ es continua y además para $u, v \in H_0^1(\Omega)$, tenemos

$$\begin{aligned}[\Phi'(u) - \Phi'(v)](u - v) &= \|u - v\|^2 \\ &\quad - \int_\Omega [f(u) - f(v)](u - v)dx \\ &\geq \|u - v\|^2.\end{aligned}$$

lo que implica que Φ es estrictamente convexa. Por otra parte,

$$\begin{aligned}\Phi(u) &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 - \int_\Omega F(u)dx \\ &= \frac{1}{2}\|u\|^2 - \int_\Omega F(u)dx \\ &\geq \frac{1}{2}\|u\|^2 \text{ (pues, } F(u) \leq 0, \forall u)\end{aligned}$$

lo que implica que Φ es coerciva. Se deduce que Φ admite un único mínimo global en $H_0^1(\Omega)$. □

1. Struwe, Michael, *Variational Methods. Applications to Nonlinear Partial Differential Equations*, Springer-Verlag, Berlin, 2008.
2. Troyanski, S. L., *Gateaux differentiable norms in Lp*, Math. Ann. 287, 221-227, 1990.
3. Krantz, Steven G., *Partial Differential Equations and Complex Analysis*, CRC Press, Inc., USA, 1992.
4. Gol'dshtein, Vladimir, *Axiomatic Theory of Sobolev Spaces*, Expo. Math. 19, 289-336, 2001.
5. Evans, Laurence C., *Partial Differential Equations*, University of California, USA, 1998.
6. Jabri, Youssef, *The Mountain Pass Theorem*, Cambridge University Press, New York, 2003.
7. Kartashov, E. M., *A new approach to the solution of Dirichlet and Neumann boundary value problems for the Laplace equations*, Thermal Engineering 57, 13, 1149-1155, 2010.
8. Chang, Jen-Shih, *Handbook of Electrostatic Processes*, Marcel Dekker, Inc., New York, 1995
9. Le Dret, Hervé, *Nonlinear Elliptic Partial Differential Equations*, Springer International Publishing AG, Switzerland, 2018.
10. Costa, David G., *An invitation to Variational Methods in Differential Equations*, Birkhäuser Boston, USA, 2007
11. Badiale, Marino, *Semilinear Elliptic Equations for Beginners*, Springer-Verlag, England, 2011.

Una Heurística de Clusterización para el Problema del Ruteo de Vehículos Multidepósito

Rósulo Hilarión Pérez Cupe[†], Luis Ernesto Flores Luyo[‡] y Rolando Raul Palomino Vildoso[‡]
Escuela Profesional de Matemática. Facultad de Ciencias. IMCA
Universidad Nacional de Ingeniería;
[†]*rperezc@uni.edu.pe* [‡]*lflores@imca.edu.pe,* [‡]*rpalominov@uni.edu.pe*

16 de noviembre del 2020; aceptado el 22 de diciembre del 2020

El problema VRP (**V**ehicle **R**outing **P**roblem) es uno de los problemas de optimización mas importantes y desafiantes en el campo de la Investigación de Operaciones, consiste en la construcción de un conjunto óptimo de rutas para una flota de vehículos que deberán satisfacer la demanda de un conjunto de clientes, el problema está clasificado como un problema combinatorio computacionalmente difícil (NP-Hard). En las aplicaciones prácticas, diferentes versiones del VRP han ido apareciendo tal como el problema MDVRP (**M**ulti **D**epot **V**ehicle **R**outing **P**roblem) [1], cuando un problema NP-Hard no puede ser resuelto de manera exacta se buscan soluciones aproximadas obtenidas mediante heurísticas. En el presente trabajo estudiamos, formulamos y resolvemos (por medio de heurísticas) el problema MDVRP, la solución aproximada se trata desde el punto de vista práctico a través de la formulación e implementación (en el lenguaje de programación JULIA 1.0.5) de las heurísticas de construcción y mejora (siendo ésta la contribución del trabajo de investigación). En cuanto a la heurística de construcción, se presentan dos propuestas de agrupamiento o clusterización basadas en la ubicación geográfica de los clientes y los almacenes, así como de su proximidad entre sí, en cuanto a las heurísticas de mejora, las estrategias del vecino más cercano y de separación fueron usados. Finalmente, se presentan los resultados y comparaciones con respecto a la solución BKS (Best Know Solution) disponible en la literatura.

Palabras clave: Heurística, NP-Hard, Clusterización, MIP.

The vehicle routing problem VRP is one of the most important and challenging optimization problems in the field of Operations Research, consists of building an optimal set of routes for a fleet of vehicles that should satisfy the demand of a set of customers, the problem is classified as a computationally hard combinatorial problem. In practical application, different needs have emerged that made it necessary to formulate extensions or variants of the VRP problem, such as the MDVRP problem (**M**ulti **D**epot **V**ehicle **R**outing **P**roblem) [1], when a computationally hard combinatorial problem like the one mentioned cannot be solved exactly, the approximate solutions obtained by heuristic methods are used. In the present work we study, formulate and solve (by means of heuristics) the MDVRP problem, the approximate solution is dealt with from the practical point of view through the formulation and implementation (in the JULIA 1.0.5 programming language) of the construction and improvement heuristics (this being the contribution of the research work). Regarding the construction heuristics, two proposals for grouping or clustering are presented based on the geographic location of customers and warehouses as well as their proximity to each other, in terms of improvement heuristics, the strategies of the closest neighbor and separation were used. Finally, the results and comparisons with respect to the best-known BKS solution (Best Know Solution) available in the literature are presented.

Keywords: Heuristic, NP-Hard, clustering, MIP.

1. Introducción

El problema MDVRP consiste en la distribución de algún producto o bien que se encuentra almacenado en depósitos y debe ser distribuido a los clientes cada uno de los cuales tiene una demanda conocida, la distribución se realiza mediante vehículos de capacidad limitada. Específicamente se tiene n depósitos en cada una de las cuales existe una flota de vehículos que trasladarán los productos a los m clientes, cada vehículo sale de un depósito visita una sola vez a cada cliente, satisface su demanda y regresa al mismo depósito de donde partió. Siendo el costo de transporte proporcional a la distancia, el objetivo será encontrar aquella ruta que minimiza la distancia total recorrida por todos los vehículos, formulándose así un problema de programación entera mixta, el problema así formulado es de naturaleza $NP-Hard$ [2], ésta es la razón por la cual encontrar una solución exacta para tamaños considerables de clientes y/o depósitos es difícil de obtener, por ello se requiere la formulación de heurísticas que se diseñarán en el desarrollo del presente trabajo.

Por ejemplo si suponemos que una empresa cervecera tiene tres depósitos principales A, B y C (tal como se muestra en la figura 1) desde los cuales debe distribuir su mercadería a 18 clientes (círculos numerados en su interior y en cuya parte superior aparece su demanda) y para ello dispone de una flota de 9 vehículos cada una con una capacidad máxima, finalmente cada depósito también posee una capacidad máxima en cuanto a productos y vehículos, se plantea así un problema de ruteo de vehículos cuya solución optimizará la operatividad de esta empresa.

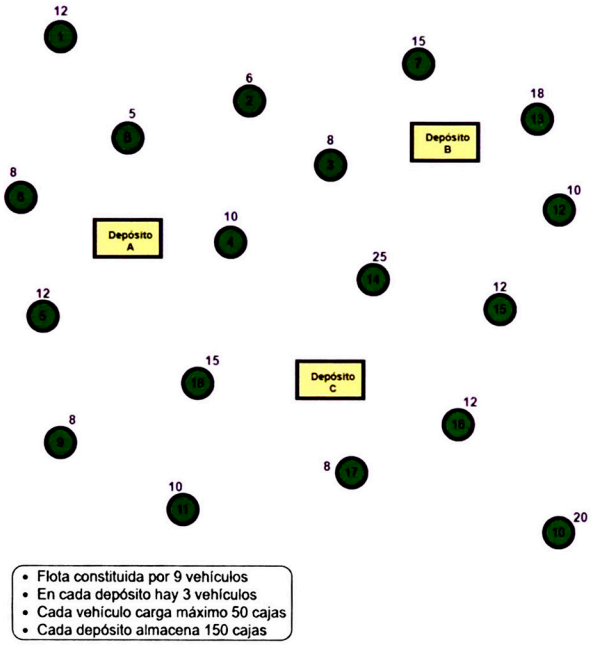


Figura 1. *Una instancia del problema del ruteo de vehículos multidepósito MDRVP con 3 depósitos y 18 clientes.*

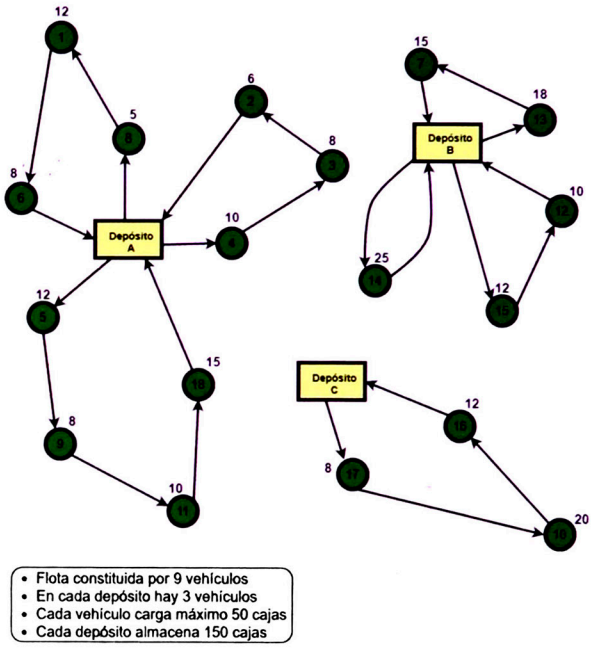


Figura 2. *Una solución factible para el problema propuesto en la figura 1. Esta solución usa 7 vehículos y tiene una longitud total de recorrido de 59.5 kilómetros y en consecuencia un costo total igual a 4760 soles.*

Como parámetros a priori se tiene información respecto a las ubicaciones de clientes y depósitos, flota de vehículos, carga máxima de cada vehículo, demanda de cada cliente, capacidad máxima de cada depósito. Consideramos la distancia entre clientes y depósitos dado en kilómetros y consideramos un costo por kilómetro igual a 80 soles (por ejemplo puede considerarse consumo de combustible, mantenimiento del vehículo y otros).

La solución exacta es posible determinarla utilizando solvers como son CPLEX versión 12.6.1 y GUROBI versión 9.1.0 o también programando el modelo en el lenguaje de programación JULIA version 1.0.5, cabe mencionar que solo es posible obtener soluciones exactas para tamaños de problema pequeños (menos de 30 clientes) en tiempos razonablemente cortos, para tamaños mayores el tiempo podría ser de orden exponencial, el análisis de complejidad de algoritmos puede encontarse en [3].

Por ejemplo en la figura 2 se tiene una solución factible para el problema planteado en la figura 1, se observa que en este caso el costo de la solución es igual a 4760 soles (de acuerdo a los datos considerados para la cuantificación del costo). Se dice que el problema está resuelto si encontramos aquella solución que minimiza el costo, al ser el problema mencionado de tipo $NP-hard$ la cantidad de soluciones factibles es de tipo combinatorio o exponencial por lo tanto una solución por exploración de todos los casos es inviable, en el presente trabajo se construirán heurísticas que permitan encontar soluciones razonablemente cercanas a la exacta.

2. Modelo Matemático

Presentamos a continuación un modelo matemático para la solución del problema planteado, consideremos para ello los siguientes conjuntos:

I : Depósitos ($i \in I$) $\qquad$ J : Clientes ($j \in J$)
K : Vehículos ($k \in K$), también denota la ruta k

Debe observarse que los conjuntos I, J y K son subconjuntos finitos de $\mathbb{N}$.

Asimismo consideremos los siguientes parámetros: (los cuales se conocen a priori)
N : cantidad de vehículos
C_{ij} : distancia entre los puntos i y j $\qquad$ $i, j \in I \cup J$
D_i : capacidad máxima del depósito i
d_j : demanda del cliente j
Q_k : capacidad máxima del vehículo (o ruta) k

A continuación se definen las variables de decisión:

$$x_{ijk} = \begin{cases} 1 & ; \text{ si el cliente } i \text{ precede inmediatamente} \\ & \quad \text{al cliente } j \text{ en la ruta } k \\ 0 & ; \text{ en otro caso} \end{cases}$$

$$z_{ij} = \begin{cases} 1 & ; \text{ si el cliente } j \text{ es asignado al depósito } i \\ 0 & ; \text{ en otro caso} \end{cases}$$

Asimismo consideramos adicionalmente las variables auxiliares U_{jk} : ($j \in J$ y $k \in K$) usados para las restricciones de eliminación de subrutas en la ruta k.

La función objetivo del modelo matemático consistirá en minimizar la distancia total de recorrido de todos los vehículos, es decir

$$\text{mín}\left\{\sum_{i\in I\cup J}\sum_{j\in I\cup J}\sum_{k\in K} C_{ij}x_{ijk}\right\}$$

sujeto a las restricciones:

$$\sum_{k\in K}\sum_{i\in I\cup J} x_{ijk} = 1 \quad \forall\, j \in J \tag{1}$$

Esta restricción nos dice que a cada cliente se le asignará una sola ruta.

$$\sum_{j\in J}\sum_{i\in I\cup J} d_j x_{ijk} \leq Q_k \quad \forall\, k \in K \tag{2}$$

Esta restricción asegura que la suma de demandas de los clientes de una determinada ruta es menor o igual a la capacidad del vehículo asignado a dicha ruta.

$$U_{lk} - U_{jk} + N x_{ijk} \leq N-1 \quad \forall\, l,j \in J \quad ; \quad \forall\, k \in K \tag{3}$$

Restricciones de eliminación de subrutas.

$$\sum_{j\in I\cup J} x_{ijk} - \sum_{j\in I\cup J} x_{jik} = 0 \quad \forall\, k \in K \quad \forall\, i \in I\cup J \tag{4}$$

Restricción referida a la conservación de flujo, para cada cliente o depósito la cantidad de arcos que ingresa al nodo es igual a la cantidad de arcos que sale de dicho nodo.

$$\sum_{i\in I}\sum_{j\in J} x_{ijk} \leq 1 \quad \forall\, k \in K \tag{5}$$

Establece que cada ruta o vehículo es usado a lo más una vez.

$$\sum_{j\in J} d_i z_{ij} \leq D_i \quad \forall\, i \in I \tag{6}$$

Esta restricción asegura que en cada depósito la suma de las demandas de los clientes abastecidos por el mencionado depósito debe ser menor que la capacidad máxima del depósito.

$$-z_{ij} + \sum_{u\in I\cup J} (x_{iuk} + x_{ujk}) \leq 1 \quad \forall\, i \in I,\ \forall\, j \in J,\ \forall\, k \in K \tag{7}$$

Esta restricción nos indica que un cliente es asignado a un depósito solo si hay ruta desde el depósito.

$$x_{ijk} \quad ; \quad z_{ij} \in \{0;1\} \quad ; \quad U_{lk} \geq 0 \tag{8}$$

Restricción de la naturaleza de las variables.

En los siguientes artículos ([5], [1] y [6]) se puede encontrar propuestas de heurísticas y metaheurísticas para el modelo planteado.

Dentro de la literatura existen dos tipos de heurísticas:

Heurísticas de Construcción: Estas heurísticas utilizan estrategias específicas para obtener una solución factible inicial, generalmente estas soluciones no son tan buenas, en el sentido de encontrarse lejos del óptimo, sin embargo gracias a las heurísticas de mejora podemos acercarnos a dicho óptimo.

Heurísticas de Mejora: Tomando la solución factible inicial obtenida por la heurística de construcción, estas heurísticas realizan pequeñas variaciones a la solución obtenida con la intención de mejorar la función objetivo.

3. Heurísticas de Clusterización

Las heurísticas propuestas en el presente trabajo construyen clústeres (grupos de clientes), cada cluster será asignados a un único depósito (un depósito puede contener mas de un cluster en general) y la demanda de cada cluster será satisfecha por la flota de vehículos existente en el depósito. Asimismo luego de obtener una solución factible inicial se proponen heurísticas de mejora para aproximarse mucho más a la mejor solución conocida (BKS)

3.1. La Heurística de clusterización por Distancias (HD)

En esta heurística se construirán tantos clústeres como depósitos existentes, para finalmente asignar a cada clúster el depósito más cercano. Cabe mencionar que en ésta propuesta de heurística no se tiene aún en cuenta las demandas de los clientes, ni la capacidad máxima de los depósitos, asi como tampoco la capacidad de cada vehículo; estos detalles serán considerados al final del proceso de clusterización y asignación de clústeres a los depósitos.

En cada paso del algoritmo se tiene una determinada cantidad de clústeres (inicialmente cada cliente constituye un cluster, por lo tanto la cantidad de clústeres al inicio es igual a la cantidad de clientes), entonces el algoritmo identifica los dos clústeres más cercanos (considerando la distancia entre dos clústeres como la distancia entre sus centroides) y une a ambos clústeres, constituyendo así un nuevo cluster de mayor tamaño (en cuanto a cantidad de nodos y demanda). El proceso se detiene cuando se ha logrado construir tantos clústeres como depósitos. Luego de considerar las siguientes notaciones: $m = |I|$ (cantidad de depósitos), $n = |J|$ (cantidad de clientes), $\mathcal{C}$:(conjunto finito de clústeres, cada elemento de $\mathcal{C}$ es un conjunto de clientes) y $d(C_i, C_j)$: distancia entre los clústeres C_i y C_j, se muestra el algoritmo

Algoritmo 1 (HD) Algoritmo de clusterización por distancias

Entrada: Clientes:(ubicación, demandas y cantidad). Depósitos:(ubicación, capacidad, cantidad). Vehículos:(capacidad y cantidad)

Salida: Conjunto de clústeres

1: $\mathcal{C} = \{\{c_1\}, \{c_2\}, \cdots, \{c_n\}\}$ (cada cliente es un cluster)
2: **mientras** $|\mathcal{C}| > m$ **hacer:**
3: Hallar dos índices i_{min} y j_{min} tales que $d(C_{i_{min}}, C_{j_{min}}) = min\{d(C_i, C_j) \,/\, C_i, C_j \in \mathcal{C}\}$
4: $C_{min} = C_{i_{min}} \cup C_{j_{min}}$
5: $\mathcal{C} = (\mathcal{C} \setminus \{C_{i_{min}}, C_{j_{min}}\}) \cup C_{min}$
6: **fin de mientras**
7: **retorna** $\mathcal{C}$

Se aplicará esta heurística a los datos mostrados (instancia) en la figura 3, donde tenemos 15 clientes enumerados desde 1 hasta 15, con sus respectivas demandas (los cuales están señalados en la parte superior de los respectivos nodos) y 5 depósitos (A , B , C , D y E), cuyas capacidades máximas también están señalados en la parte superior de cada depósito. También se observa que en la instancia de prueba se dispone de una flota de 6 vehículos $V_1, V_2, \cdots, V_6$ con sus respectivas capacidades máximas de carga (flota heterogénea en este caso).

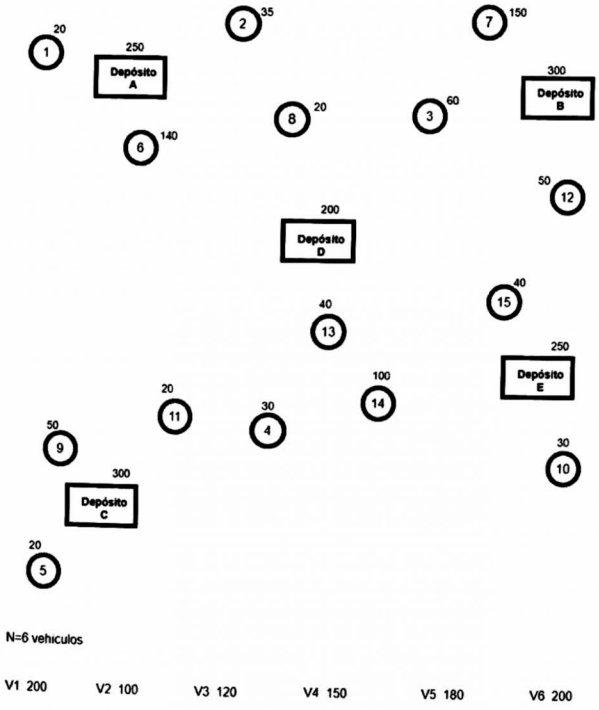


Figura 3. *Instancia de prueba a la cual se le aplicará la heurística* **HD**.

Como resultado de la aplicación del algoritmo a la instancia mencionada se obtuvieron 5 clústeres (C_1 , C_2 , C_3 , C_4 y C_5) uno por cada depósito, tal como se muestra en la figura 4.

Esta heurística presenta un inconveniente, si bien la clusterización de los clientes y depósitos es correcta, al no tener control sobre la capacidad total de cada clúster obtenido y las capacidades del vehículo o vehículos a asignar podría tornarse complicado o incluso imposible de asignar rutas. Por ejemplo según las demandas totales de los clústeres C_1 y C_2 y teniendo en cuenta las capacidades máximas de los vehículos disponibles sería necesario asignar dos vehículos a cada uno de dichos clústeres, usando así 4 vehículos de los 6 en total quedando 3 clústeres por asignar y solo dos vehículos disponibles los cuales de ninguna manera lograrán satisfacer la demanda de los otros clústeres.

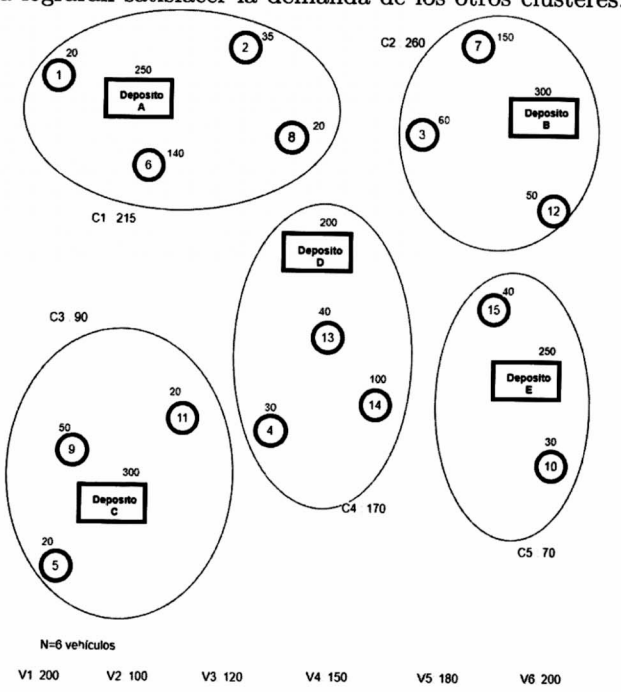


Figura 4. *Se obtuvieron 5 clústeres (igual al número de depósitos) C_1, C_2, C_3, C_4 y C_5 se muestran las demandas totales de cada clúster formado.*

Debe notarse que en cada paso del algoritmo la cardinalidad de $\mathcal{C}$ (que representa a la cantidad de clústeres) disminuye en una unidad garantizando de este modo la finalización del algoritmo. Sin embargo el algoritmo tiene algunas deficiencias, los cuales pasamos a detallar:

- El hecho de construir tantos clústeres como depósitos constituye una pérdida de generalidad, en casos donde hay pocos depósitos y muchos clientes podría definir algún cluster con muchos clientes y en consecuencia una demanda grande en comparación con la capacidad del vehículo.

- El haber construido clústeres sin considerar sus demandas podría derivar en la no existencia de un vehículo que satisfaga la demanda de algún cluster, tal como ocurre en la clusterización obtenida para la instancia mostrada en la figura 3

El último ejemplo mostrado en las figuras 3 y 4 demuestran que en general el algoritmo propuesto no funcionará, sin embargo tomando como base la idea de la clusterización, en la siguiente subsección, construimos los clústeres ya pensando también en las futuras rutas, es decir ya asignando (de manera eficiente) los vehículos disponibles, de tal manera que al final del proceso ya se tienen las rutas con sus respectivos vehículos asignados.

3.2. La heurística de clusterización por distancias y demandas (HDD)

Para evitar lo ocurrido con la heurística **HD**, construiremos los clústeres considerando no solo las distancias (cercanías) sino también las demandas y la asignación de vehículos a cada clúster, esta heurística a la que llamaremos **HDD** está compuesto de tres partes: Clusterización, Asignación óptima y TSP, cada una de ellas se describen en las siguientes subsecciones.

3.2.1. Clusterización

Para esta parte de la heurística definimos algunos términos y notaciones que serán utilizados en la heurística.

1. Se definen los conjuntos $\mathcal{C}_{temp}$: que contiene a los clústeres temporales aún en proceso de depuración y construcción y $\mathcal{C}_{def_i}$: que contiene a los clústeres definitivos.

2. Se dice que la construcción de clústeres temporales **colapsa** cuando se elige dos clústeres más cercanos (cada uno con vehículo asignado) y NO existe vehículo que pueda atender la demanda de la unión de dichos clústeres.

3. Si $V_1, V_2, \cdots, V_t$ son vehículos disponibles para ser asignados al clúster C (cuya demanda total es d_C) y $Q_{V_1}, Q_{V_2}, \cdots, Q_{V_t}$ son las cargas máximas de los respectivos vehículos. El vehículo V_k $(k = 1, 2, \cdots, t)$ se llama **mejor vehículo** si y solo si

$$Q_{V_k} - d_C = \min_{1 \leq i \leq t} \{Q_{V_i} - d_C \ / \ Q_{V_i} > d_C\}$$

   Por ejemplo si un clúster C tiene demanda total $d_C = 190$ y se tienen $t = 4$ vehículos disponibles de capacidades 200, 160, 250 y 300, el **mejor vehículo** es el de capacidad 200.

4. Dados dos clústeres con vehículos asignados (C_1, V_1) y (C_2, V_2) y demandas totales conocidas, diremos que (C_1, V_1) es **más saturado** que (C_2, V_2) si $Q_{V_1} - d_{C_1} \leq Q_{V_2} - d_{C_2}$.

5. El control de los vehículos asignados se lleva a cabo a través de la función $v : \mathcal{C} \to \{0, 1\}$, donde $\mathcal{C}$ es el conjunto actual de clústeres y para cada $C \in \mathcal{C}$

$$v(C) = \begin{cases} 1 & ; \quad C \text{ tiene vehículo asignado} \\ 0 & ; \quad \text{en caso contrario} \end{cases}$$

Con respecto a la heurística anterior han sido mejorados varios aspectos, los cuales mencionamos a continuación:

- Los clústeres serán construidos teniendo en cuenta las distancias pero también la posibilidad de asignárseles el **mejor vehículo**.

- La idea principal en la presente heurística es seleccionar los dos clústeres más cercanos, intentar unirlos en un solo clúster y asignarle el **mejor vehículo**. Este proceso repetitivo inevitablemente **colapsa** en la imposibilidad de unir dos clústeres más cercanos con vehículos ya asignados. Al ocurrir este caso se envía el clúster **más saturado** al conjunto $\mathcal{C}_{def_i}$ dejando al otro clúster en $\mathcal{C}_{temp}$, este proceso continúa hasta que solo queda un clúster en el conjunto $\mathcal{C}_{temp}$.

- Dependiendo del caso, un vehículo que previamente fue asignado a un clúster podría ser liberado posteriormente. Por ejemplo si el algoritmo selecciona dos clústeres C_i y C_j(más cercanos) cuyas demandas totales hasta el momento son 30 y 80 respectivamente y tienen asignados dos vehículos cuyas capacidades son 100 y 150 respectivamente, entonces los clústeres mencionados se unirán para constituir un solo clúster de capacidad $30 + 80 = 110$ al cual se le asigna el vehículo de capacidad 150 quedando libre el vehículo de capacidad 100 para que sea utilizado en otro cluster.

El algoritmo selecciona en cada paso los dos clústeres más cercanos C_i y C_j, y dependiendo de los vehículos asignados a dichos clústeres se analizan los siguientes tres casos:

1. **Ambos no tienen asignado un vehículo:** (esto ocurre en los primeros pasos), se escoge el **mejor vehículo** para cada uno o si es posible se une a ambos en un solo cluster y se le asigna el **mejor vehículo** disponible siempre que exista.

2. **Ambos tienen asignado un vehículo:** en este caso se intenta unir ambos clústeres en uno solo con el mejor vehículo escogido entre los dos vehículos que estan siendo usados o algún otro disponible. Cuando no es posible realizar la unión ocurre el **colapso**.

3. **Solo uno de los clústeres tiene un vehículo asignado:** en este caso se intenta aglutinar en el clúster con vehículo todo el contenido del otro clúster, en caso no fuera posible se busca el mejor vehículo disponible para el clúster sin vehículo.

El algoritmo termina cuando hay un solo clúster en el conjunto $\mathcal{C}_{temp}$ el cual pasa a formar parte del conjunto $\mathcal{C}_{def_i}$ quedando el conjunto de clústeres temporales vacío $(\mathcal{C}_{temp} = \phi)$, el cumplimiento de ésta condición está garantizado dado que en cada paso se eligen los dos clústeres más cercanos y se unen para formar otro cluster de mayor demanda y debido a la capacidad limitada de los vehículos asignados inevitablemente se llegará al **colapso** lo cual obliga a eliminar un clúster de $\mathcal{C}_{temp}$ y enviarlo al conjunto $\mathcal{C}_{def_i}$ permitiendo así que la cantidad de clústeres en $\mathcal{C}_{temp}$ disminuya sistemáticamente. Finalmente los clústeres obtenidos luego del proceso anterior deberán ser asignados de manera óptima a los depósitos para construir las rutas. Este procedimiento se detalla en la siguiente subsección.

Con todas las consideraciones anteriores mostramos el algoritmo:

Algoritmo 2 (HDD) Algoritmo de clusterización por distancias y demandas

Entrada: Clientes:(ubicación, demandas y cantidad). Depósitos:(ubicación, capacidad, cantidad). Vehículos:(capacidad y cantidad)

Salida: Conjunto de clústeres $\mathcal{C}$

$\mathcal{C}_{temp} = \{\{c_1\}, \{c_2\}, \cdots, \{c_n\}\}$
$\mathcal{C}_{defi} = \{\phi\}$
$v(C) = 0 \ \forall C \in \mathcal{C}_{temp}$
mientras $|\mathcal{C}_{temp}| > 1$ **hacer:**
  Hallar los índices i_{min} y j_{min} tales que $d(C_{i_{min}}, C_{j_{min}}) = min\{d(C_i, C_j) \ / \ C_i \ , C_j \in \mathcal{C}_{temp}\}$
  si $(v(C_{i_{min}}) + v(C_{j_{min}}) = 0)$ **entonces**
    Juntarlos en un solo clúster y asignarle el mejor vehículo en caso se pueda o asignarle a cada clúster un mejor vehículo
  sino
    si $(v(C_{i_{min}}) + v(C_{j_{min}}) = 2)$ **entonces**
      si ocurre **colapso entonces**
        $C =$ más saturado entre $C_{i_{min}}$ o $C_{j_{min}}$
        $\mathcal{C}_{temp} = \mathcal{C}_{temp} \setminus \{C\}$
        $\mathcal{C}_{defi} = \mathcal{C}_{defi} \cup \{C\}$
      sino
        $C = C_{i_{min}} \cup C_{j_{min}}$
        Asignar el mejor vehículo disponible a C y liberar el otro vehículo
        $\mathcal{C}_{temp} = (\mathcal{C}_{temp} \setminus \{C_{i_{min}}, C_{j_{min}}\}) \cup \{C\}$
      fin de si
    sino
      si $(v(C_{i_{min}}) + v(C_{j_{min}}) = 1)$ **entonces**
        $C_1 =$ clúster con vehículo asignado
        $C_0 =$ clúster sin vehículo asignado
        si C_0 puede aglutinarse en C_1 **entonces**
          $C_1 = C_1 \cup C_0$
          $\mathcal{C}_{temp} = \mathcal{C}_{temp} \setminus \{C_0\}$
        sino
          Asignar el mejor vehículo disponible a C_0
        fin de si
      fin de si
    fin de si
  fin de si
fin de mientras
$\mathcal{C} = \mathcal{C}_{temp} \cup \mathcal{C}_{defi}$
retorna $\mathcal{C}$

3.2.2. Asignación óptima de clústeres a depósitos

Luego de obtener los clústeres con vehículos ya asignados y teniendo toda la información respecto a los depósitos, a continuación debemos asignar los clústeres obtenidos a los depósitos. Eventualmente algunos depósitos podrían quedar sin ningún clúster asignado o un depósito podría albergar a más de un clúster siempre que su capacidad máxima no haya sido alcanzada. Para resolver éste subproblema, planteamos un problema de optimización binaria (dado que la variable de decisión solo puede tomar el valor de 0 o 1), para ello consideremos los siguientes parámetros:

p : cantidad de clústeres obtenidos
m : cantidad de depósitos
$i = 1; 2; \cdots; m$ (índice de depósitos)
$j = 1; 2; \cdots; p$ (índice de clústeres)
M_{ij} : representa la distancia del depósito i al cluster j
d_{C_j} : es la demanda total del clúster j
D_i : es la capacidad máxima del depósito i
NVD : es la cantidad de vehículos en cada depósito
Consideramos la variable de decisión:

$$x_{ij} = \begin{cases} 1 & ; \quad \text{si el cluster } j \text{ es asignado al depósito } i \\ 0 & ; \quad \text{en otro caso} \end{cases}$$

La función objetivo para este subproblema consistirá en minimizar la distancia entre clústeres asignados y los depósitos, es decir,

$$\text{mín} \left\{ \sum_{j=1}^{p} \sum_{i=1}^{m} M_{ij} x_{ij} \right\}$$

sujeta a las restricciones:

$$\sum_{j=1}^{p} d_{C_j} x_{ij} \leq D_i \quad \forall \, i = 1; 2; \cdots; m \tag{9}$$

Esta restricción asegura que la suma de demandas de los clústeres asignados al depósito i no superan su capacidad.

$$\sum_{i=1}^{m} x_{ij} = 1 \quad \forall \, j = 1; 2; \cdots; p \tag{10}$$

Esta restricción garantiza que cada clúster es asignado a un solo depósito.

$$\sum_{j=1}^{p} x_{ij} \leq NVD \quad \forall \, i = 1; 2; \cdots; m \tag{11}$$

Esta restricción garantiza que la cantidad de clústeres asignados a cada depósito es menor o igual a la cantidad de vehículos en dicho depósito.

$$x_{ij} \in \{0; 1\} \tag{12}$$

Restricción que expresa el tipo de variable.

Esta parte de la heurística se implementa en el lenguaje de programación JULIA 1.0.5, dado que en general el tamaño del problema (expresado en términos de cantidad de clústeres y depósitos) es relativamente pequeña, respecto al tamaño inicial del problema.

Para la misma instancia de prueba mostrada en la figura 3, se puede observar en este caso la obtención de 6 clusteres (color rojo) con el mejor vehículo ya asignado vea la figura (5). A continuación la solución del problema de asignación óptima de clústeres a depósitos permite asignar a cada clúster un único depósito, obteniendo clústeres de mayor tamaño (a los que llamaremos superclústeres) que incluyen a los clústeres asignados a un determinado depósito y también al depósito. Estos superclusteres (que agrupan a varios clústeres en torno

a un mismo depósito asignado) están representados en la figura (5) de color azul, existen tantos superclústeres como depósitos.

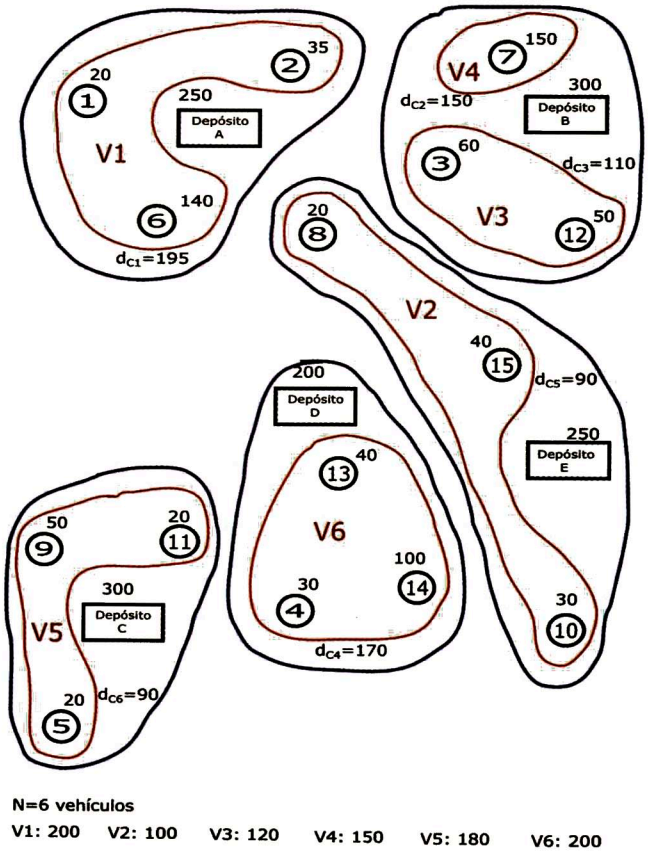


Figura 5. *Los clústeres de color rojo son los 6 originales de acuerdo a las distancias y demandas además cada clúster ya tiene asignado un vehículo, mientras que los superclústeres de color azul en general aglutinan a varios clústeres y sus depósitos asignados, eventualmente algún depósito podría quedar sin clústeres asignados.*

3.2.3. TSP para definir las rutas

Luego de culminar el proceso de clusterización y la asignación de cada uno de los clústeres a los depósitos, se construyen las rutas resolviendo para cada cluster y su depósito asignado un problema del agente viajero TSP, definiendo la ruta de menor distancia, obteniendo así un conjunto de rutas que constituye una solución inicial factible. La heurística está diseñada para ser aplicada a un problema MDVRP con flota mixta y con restricciones de capacidad máxima para los depósitos. Sin embargo con el objetivo de realizar las pruebas para las instancias definidas en el marco del presente trabajo `https://github.com/fboliveira/MDVRP-Instances/` se considera la misma cantidad de vehículos en cada depósito y todos los vehículos con la misma capacidad es decir flota homogénea. Resumimos la heurística **HDD** simbólicamente en la expresión:

$$\text{HDD} = \text{Clusterización} + \text{Asignación Óptima} + \text{TSP}$$

Esta heurística ha sido implementada en el lenguaje de programación JULIA 1.0.5 y se usará una instancia creada manualmente de tamaño pequeño (15 clientes y 5 depósitos) con el fin de obtener también la solución exacta para efectos de comparación, se muestran además las ubicaciones iniciales de los clientes y depósitos, el resultado de la clusterización por distancias y demandas y la solución exacta para la instancia creada llamada **DataPrueba**

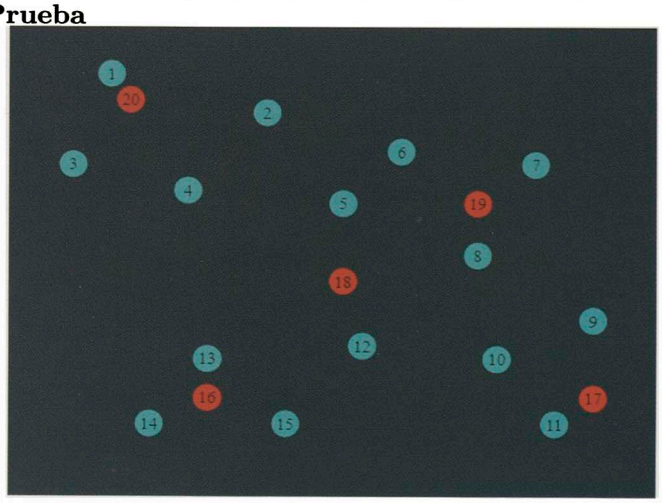


Figura 6. *Nube de puntos correspondiente a la instancia manual* **DataPrueba** *con 5 depósitos (de color rojo y enumerados de 16 a 20), 15 clientes (de color turquesa y enumerados de 1 a 15) y dos vehículos cuya carga máxima es 60 en cada depósito.*

Al aplicar la heurística por distancias y demandas (HDD) desarrollada en esta sección a la instancia **DataPrueba** se obtiene como resultado la solución que se muestra en la siguiente figura.

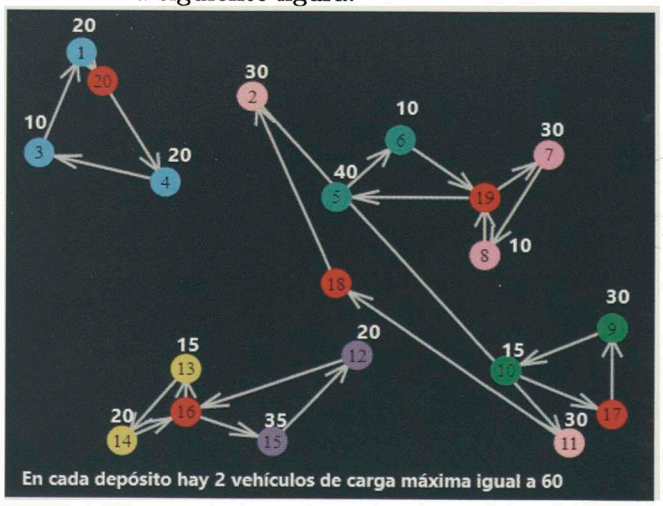


Figura 7. *Resultados obtenidos por la aplicación de la heurística HDD a la instancia* **DataPrueba**, *los conjuntos de nodos con el mismo color representan nodos del mismo clúster cuyas demandas serán satisfechas por un vehículo de la flota, los depósitos estan representados con el color rojo, el número mostrado en la parte superior de cada nodo cliente representa su demanda. La función objetivo que resulta de la aplicación de la heurística proporciona el valor* **165.15** *utilizando 7 vehículos (de la flota de 10 vehículos).*

Se muestra a continuación para la misma instancia la solución exacta (por implementación del modelo matemático en JULIA 1.0.5) se observa similitud en alguno de los clústeres obtenidos entre la solución por heurística

y la solución exacta.

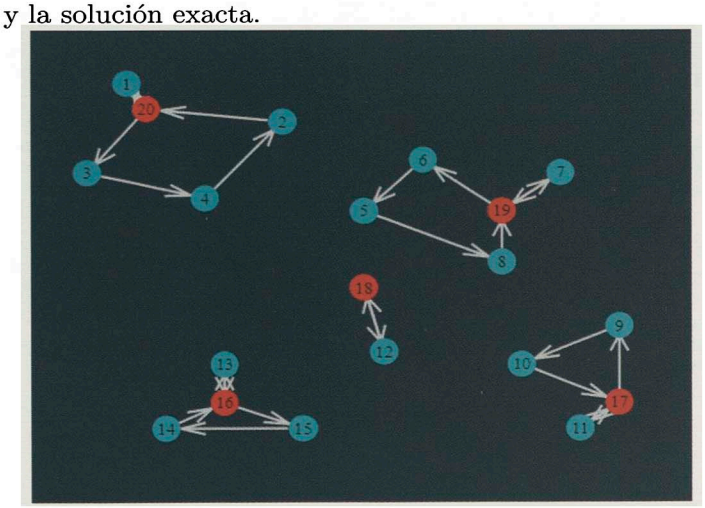


Figura 8. *Solución exacta para la misma instancia, los depósitos estan representados de color rojo y sus respectivos clusteres de color turquesa alrededor del depósito, el valor objetivo óptimo es* **116.70** *en este caso se usan 9 vehículos (del total de 10 vehículos).*

Finalmente elaboramos una tabla comparativa con el desempeño de la heurística **HDD** para la instancia creada **DataPrueba** la cual tiene $m = 5$ depósitos, $n = 15$ clientes, así mismo otras características de la instancia son la existencia de 2 vehículos por depósito, es decir $TV = 10$ (tamaño de flota) y para ésta instancia en particular (tamaño pequeño) es posible hallar la solución exacta, la cual es 116,70 hallada en 6.33 segundos. Todos estos datos y resultados obtenidos por la ejecución de las heurísticas y el modelo exacto son resumidas en la tabla 1

Método	costo	Nro de Vehículos	tiempo(seg)
solución exacta	116.70	9	6.33
Heurística HDD	165.15	7	1.09

Cuadro 1. Resultados obtenidos para la solución exacta.

3.3. Heurísticas de Mejora (HM)

Luego de obtener una solución inicial factible mediante la heurística de construcción descrita anteriormente, se aplicará a la solución obtenida otra heurística llamada de mejora, dentro de éstas se han considerado las siguientes estrategias:

- **Intercambio de nodos** Consiste en trabajar en clústeres cercanos e intercambiar nodos con el objetivo de mejorar la función objetivo
- **Partición de un clúster** Luego de detectar un clúster con mucha variabilidad o irregularidad (por ejemplo es geométricamente muy grande o cruza a otros clústeres) se divide el cluster en dos o mas partes dentro de las condiciones de factibilidad.

La aplicación sucesiva de estas dos heurísticas permiten mejorar la función objetivo hasta determinado límite, se ha registrado en las tablas la cantidad de mejoras realizadas y el tiempo adicional utilizado. Para una descripción algorítmica del proceso de mejora consideramos las siguientes observaciones:

- $\mathcal{S}$ es el conjunto de soluciones factibles para el problema de optimización planteado, es decir,

  $\mathcal{S} = \{s \ / \ s \text{ es factible para el problema planteado}\}$

- La función $f : S \longrightarrow \mathbb{R}^+$ definida mediante $f(s) =$ valor objetivo asociada a la solución s
- Consideramos ya implementado el procedimiento **Mejora** que recibe una solución factible $s \in S$ y retorna una solución mejorada s_m (también factible), en el sentido siguiente: luego de aplicar

  $$s_m = \textbf{Mejora}(s)$$

  se cumple

  $$f(s_m) \leq f(s)$$

Donde se verifica $f(s_m) = f(s)$ cuando ya no se puede mejorar la solución s. En caso que $f(s_m) < f(s)$ se ha logrado mejorar la función objetivo, describimos a continuación el algoritmo de mejoras sucesivas

Algoritmo 3 (**HM**) Algoritmo de Mejoras sucesivas

Entrada: s_0: Solución inicial dada por la heurística de construcción **HDD**
Salida: s_1: Mejor solución

1: $s_1 = \textbf{Mejora}(s_0)$
2: **mientras** $f(s_1) \neq f(s_0)$ **hacer:**
3: $\quad s_0 = s_1$
4: $\quad s_1 = \textbf{Mejora}(s_0)$
5: **fin de mientras**
6: **retorna** s_1

En la siguiente figura se muestra la secuencia de mejoras sucesivas aplicadas a la instancia p01, cuya solución factible inicial se muestra en la siguiente figura, en la que se señala el cluster de color amarillo que será modificado localmente.

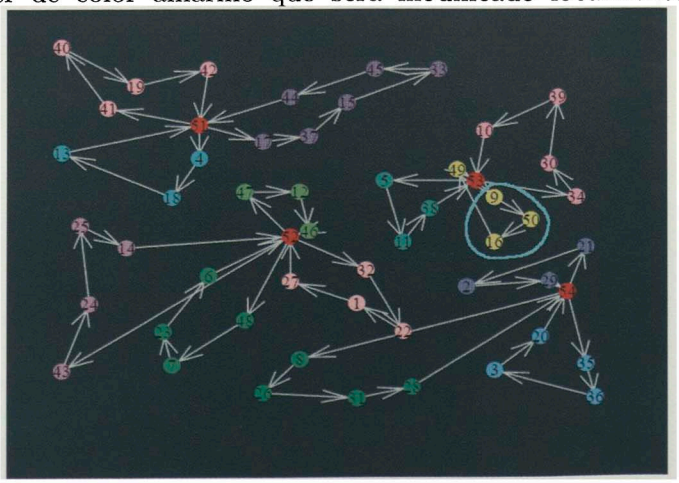


Figura 9. *Instancia p01 con 50 clientes, 4 depósitos, 4 vehículos por depósito c/u con capacidad 80. Tiempo=0.86 seg y Valor objetivo=643.699.*

A continuación se muestra una aplicación de la heurística de mejora en el cluster señalado, se observa que

los clientes 16 y 50 que originalmente pertenecían al cluster de color amarillo fueron reasignados a otros clústeres cercanos, mejorando de esta manera la función objetivo.

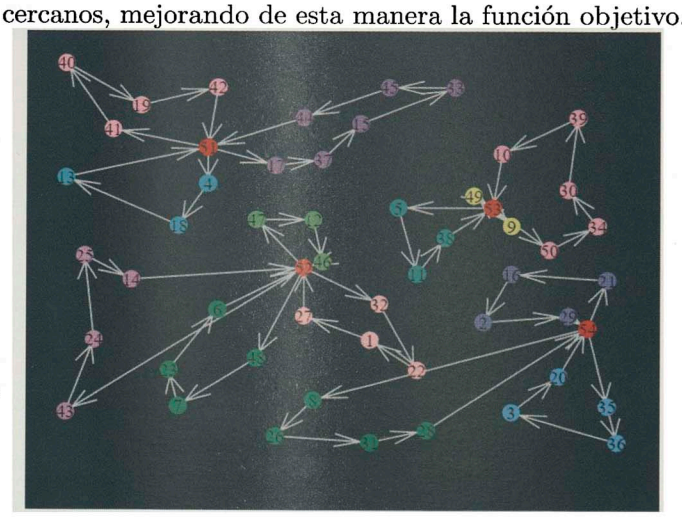


Figura 10. *Valor objetivo mejorado: 633.237 .*

Al aplicar sucesivamente las heurísticas de mejora, hasta que la función objetivo ya no se pueda mejorar, se muestra la secuencia de mejoras.

Mejora	Valor objetivo
0	643.69
1	633.23
2	627.78
3	624.55
4	614.18
5	610.92
6	603.78
7	588.49
8	588.49

Cuadro 2. Secuencia de mejoras aplicada a la instancia p01, desde la solución inicial s_0 obtenida por la heurística de construcción HDD (cuyo valor fue 643.69) hasta la solución mejorada s_1 (por aplicación sucesiva de las heurísticas de mejora) cuyo mejor valor es 588.49

Obteniendo finalmente la siguiente configuración, llamado también mejor solución para la instancia p01

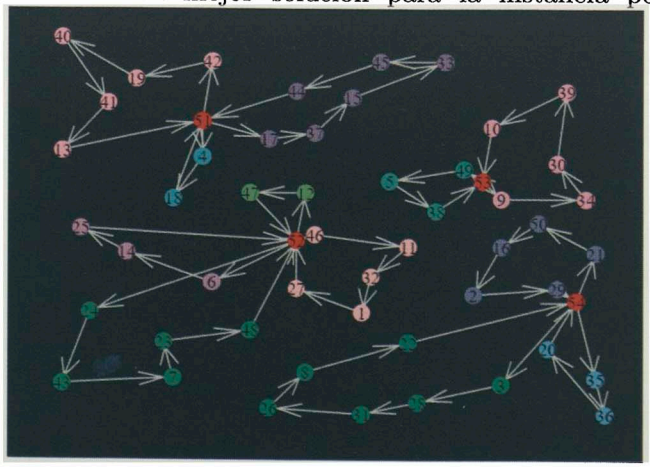


Figura 11. *Valor objetivo mejorado al máximo: 588.495.*

Para la instancia p01 en particular se ha hallado la solución exacta (al 27.2% de gap y limitada a casi 5 horas), obteniendo la siguiente configuración.

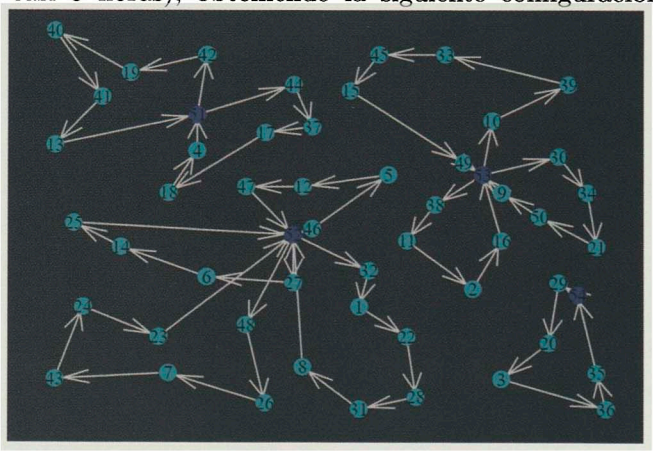


Figura 12. *Instancia p01. Tiempo=17881 seg=4.96 horas, Mejor Valor objetivo=584.43 (27.2 % de gap), Mejor conocida=576.87.*

En la siguiente sección se analiza cada una de las instancias existentes en la literatura (desde p01 hasta p17)

4. Resultados y conclusiones

4.1. Resultados para la heurística HDD

Se muestra a continuación los resultados obtenidos al aplicar la heurística a diferentes instancias existentes en la literatura.

Instancias		MEJOR SOLUC.	Soluc. Inicial por **HDD**		Solución Mejorada **HDD+HM**	
Nro	N	BKS	dist.	t_{ini}	dist.	t_{acum}
p01	50	576.87	643.69	0.85	588.49	10.54
p02	50	473.53	630.10	0.86	585.87	14.90
p03	75	641.19	720.27	0.93	689.38	10.74
p04	100	1001.59	1232.25	0.70	1145.16	45.16
p05	100	750.03	996.45	0.74	831.44	139.86
p06	100	876.50	1119.44	0.73	1017.71	27.29
p07	100	881.97	1137.23	0.71	997.12	54.32
p08	249	4372.78	5727.57	0.90	5181.44	564.37
p09	249	3858.66	4647.73	0.90	4346.48	391.33
p10	249	3631.11	4602.74	0.92	4202.72	241.26
p11	249	3546.06	4393.01	0.91	4154.11	790.99
p12	80	1318.95	2707.42	0.79	1770.32	131.23
p13	80	1318.95	2707.42	0.77	1770.32	78.14
p14	80	1360.12	2707.42	0.78	1770.32	92.28
p15	160	2505.42	5008.94	0.91	4430.36	108.80
p16	160	2572.23	5008.94	0.89	4430.36	89.24
p17	160	2709.09	5008.94	0.88	4430.36	44.93

Cuadro 3. Solución factible inicial obtenida por la heurística de construcción **HDD** basada en clusterización y la aplicación sucesiva de las heurísticas de mejora **HM**.

En la tabla anterior se muestran las características de cada una de las 17 instancias(existen 23 en la literatura), como son cantidad de clientes(N), cantidad de depósitos(M), cantidad de vehículos en cada depósito(K) y la carga máxima de cada vehículo(Q). Asimismo para cada

instancia se consigna la mejor solución conocida hasta el momento o BKS (Best Know Solution). Para cada instancia y respecto a la solución inicial tenemos registrada la solución obtenida por aplicación de la heurística en cuanto a distancia total del ruteo y tiempo de cómputo. Respecto a la solución mejorada tenemos registrado la mejor distancia obtenida, el tiempo de cómputo acumulado que incluye al tiempo empleado en hallar la solución inicial. En la tabla anterior se muestra la mejor solución conocida (BKS) hasta el momento, la distancia total obtenida por la solución, el tiempo que le toma a la heurística de construcción y el tiempo total que incluye el tiempo utilizado para hallar la solución inicial. El desempeño promedio se muestra en la siguiente gráfica donde se observa la evolución de la heurística desde la solución inicial hasta la mejora sucesiva.

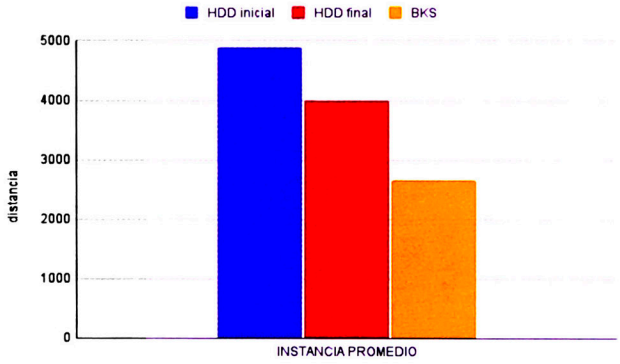


Figura 13. *Desempeño promedio de la heurística de clusterización por distancias y demandas* **HDD**.

La gráfica muestra la solución inicial promedio (azul) la mejora promedio (rojo) y la mejor solución conocida (amarillo)

4.2. Conclusiones y recomendaciones

La heurística **HD** no funciona en general sin embargo constituye la base para la elaboración de la heurística **HDD**, en promedio la heurística **HDD** (incluida su mejora) aún se puede perfeccionar dado que la brecha respecto al BKS es posible disminuirla, escencialmente el trabajo de mejora se debe enfocar en el proceso de clusterización dado que las otras dos componentes de la heurística (asignación óptima y TSP ya son óptimos). La solución exacta solo se puede hallar para tamaños muy pequeños ($n \leq 30$ clientes).

Las heurísticas de mejora utilizadas en el presente trabajo **vecino más cercano** y **spliting** permiten mejorar la función objetivo a partir de una solución inicial pero consumen un tiempo adicional.

El lenguaje de programación JULIA (de libre distribución) es una herramienta fundamental en la implementación de las heurísticas del problema de investigación planteado en el presente trabajo. Aún es posible mejorar el proceso de clusterización con otras estrategias mas específicas como $k-$means, quedando aún pendiente la propuesta de Metaheurísticas como son los algoritmos genéticos o la búsqueda tabú.

Finalmente también está pendiente la propuesta de otras estrategias de clusterización que mejoran los resultados obtenidos, estas serán presentadas en un posterior trabajo.

1. Surekha, Paneerselvam and Sumathi, Sai,World Applied Programming,1,3,118–131,2011.

2. J.K. Lenstra and A.H.G. Kan, Networks, 11, 2, 221–227, 1981, Wiley Online Library.

3. Cormen, Thomas H and Leiserson, Charles E and Rivest, Ronald L and Stein, Clifford, *Introduction to algorithms*, 2009, MIT press.

4. Pichpibul, Tantikorn and Kawtummachai, Ruengsak, ScienceAsia, 38, 3,307–318, 2012.

5. Shi, Yanjun and Lv, Lingling and Hu, Fanyi and Han, Qiaomei, Applied Sciences, 10, 7, 2403, 2020, Multidisciplinary Digital Publishing Institute.

6. Stodola, Petr, Algorithms, 11, 5, 74, 2018, Multidisciplinary Digital Publishing Institute.

Ceros de una familia de funciones enteras generadas por la función zeta de Riemann

Manuel Toribio Cangana[†] y Oswaldo Velásquez Castañón[‡]
Facultad de Ciencias.
Universidad Nacional de Ingeniería;
[†]*mtoribio@uni.edu.pe*
[‡]*ovelasquez@uni.edu.pe*

Recibido el 04 de noviembre del 2020; aceptado el 28 de diciembre del 2020

En este trabajo refrendamos que los ceros no triviales de la función zeta de Riemann están en la banda crítica, y al analizar las sumas parciales de la serie que genera la función zeta, vemos que son funciones casi-periódicas en el sentido de Bohr, así la parte real de la nube de ceros de estas sumas parciales estan acotadas y son densas en cada intervalo $[a_N, b_N]$ donde $a_N \to -\infty$ (Montgomery 1983) y $b_N \to 1$ (Velásquez Castañón 2009). Calculamos aquí buenas aproximaciones para a_N y b_N, y en cada rectángulo con altura significativa, mostramos la distribución de los ceros, y se calcula el número de ellos en esta región con una buena aproximación con respecto a resultados analíticos demostrados por Gonek y Ledoan.

Palabras Claves: Función zeta de Riemann. Ceros de funciones enteras. Aplicación del principio del argumento.

In this work, we endorse that the non-trivial zeros of the Riemann zeta function are in the critical band, and when we analyze the partial sums of the serie generated by the zeta function, we see that they are quasi-periodic functions in the Bohr sense, so the real part of the zero cloud of these partial sums is bounded and dense in each interval $[a_N, b_N]$ where $a_N \to -\infty$ (Montgomery 1983) and $b_N \to 1$ (Velásquez Castañón 2009). Here we calculate good approximations for a_N and b_N, and in each rectangle with significant height we show the distribution of the zeros, and the number of them in this region is calculated with a good approximation with respect to the analytical results demonstrated by Gonek and Ledoan.

Keywords: Riemann zeta function. Zeros of integer functions. Application of the principle of the argument.

1. Introducción

El objeto de estudio de esta investigación es la *función zeta de Riemann*, función de variable compleja definida como la prolongación analítica a todo el plano complejo de

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s};\ s \in \mathbb{C}\,,\ \Re(s) > 1 \qquad (1)$$

como una función meromorfa con un único polo simple en 1 y residuo también 1. Una relación trascendente de esta función es

$$\zeta(s) = 2(2\pi)^{s-1}\Gamma(1-s)\zeta(1-s)\text{sen}\left(\frac{\pi s}{2}\right) \qquad (2)$$

conocida como la *ecuación funcional* de la función zeta de Riemann, y el resultado de Euler ($s \in \mathbb{R}$) generalizado por Riemann.

$$\zeta(s) = \prod_{p} \left(1 - p^{-s}\right)^{-1} \quad \Re(s) > 1 \qquad (3)$$

donde p recorre el conjunto de los primos positivos. A partir de la ecuación (3) se sabe que esta función no tiene ceros en $\Re(s) > 1$. Para $\Re(s) < 0$, el teorema de reflexión de Shwarz garantiza que la función Γ no tiene ceros en todo el plano complejo, y como $\zeta(1-s) \neq 0$ se tiene que s es un cero de ζ en este semiplano si y solo si $\text{sen}\left(\frac{\pi s}{2}\right) = 0$, obteniendo así, los ceros triviales de la función zeta de Riemann $-2, -4, -6, -8, -10, \cdots$. También de (2), vemos que los posibles ceros de la función ζ en la *banda crítica* $0 \leq \Re(s) \leq 1$ guardan una simetría respecto a la *recta crítica* $\Re(s) = 1/2$, se sabe además que para todo t real $\zeta(1+it) \neq 0$, en este contexto, se genera la famosa conjetura de Riemann.

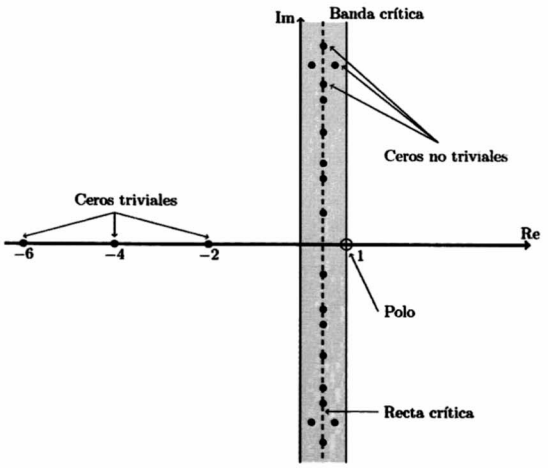


Figura 1. *La banda crítica de la función zeta*

"Los ceros no triviales de la función zeta están en la recta crítica $\Re(s) = 1/2$"

Con la tecnología de nuestros días en abril del 2020 David J. Platt de la universidad de Bristol (Inglaterra) y

Timothy S. Trudgian (Australia) [1] han demostrado que la hipótesis de Riemann es verdadera hasta una altura de 3×10^{12}, esto es, si

$$\zeta(\sigma + it) = 0 \text{ y } 0 < t < 3 \times 10^{12}, \text{ entonces } \sigma = 1/2$$

Sobre la cantidad de ceros hasta una altura prefijada, en julio del 2021 E. Hasanalizade, Q. Shen y P. Jie Wong [2] han estimado que la cantidad de ceros no triviales de la función zeta de Riemann $N(T)$ en la banda crítica de altura T, satisface la siguiente estimativa

$$\left|N(T) - \frac{T}{2\pi}\log\left(\frac{T}{2\pi e}\right)\right| \leq a\log T + b\log\log T + c$$

para $a = 0{,}1038$, $b = 0{,}2573$ y $c = 9{,}3675$

2. Consideraciones teóricas

2.1. Sobre los ceros de $\zeta_N(s)$

En este trabajo se hace un estudio de esta función ζ, a partir de la información que nos brinda las sumas parciales $\zeta_N(s) = \sum_{n=1}^{N} n^{-s}$. Veremos que los ceros de esta familia de funciones estan ubicados en bandas verticales de ancho finito.

Sobre la búsqueda de los ceros de ζ_N en el semiplano $\Re(s) > 1$, se sabe que en 1948 Paul Turán demostró que ζ_N no tiene ceros en $\Re(s) > 1$ para $N = 1, 2, 3, 4$ y 5 [3], en 1966 Robert Spira demostró que ζ_N no tiene ceros en $\Re(s) > 1$ para $N = 6, 7, 8$ y 9 [4], dos años mas tarde el mismo R. Spira demuestra que ζ_N si tiene ceros en el semiplano $\Re(s) > 1$ para $N = 19;\ 22, 23, 24, 25, 26, 27;\ 29, 30, 31, \cdots, 49$ y 50, en 1980 W.R. Monach demuestra que $\zeta_N(s) = 0$ si tiene solución en $\Re(s) > 1$ para $N \geq 51$ [5]. Cerrando este enfoque en el año 2016 D.J. Platt y T.S. Trudgian [6] demuestran que para

$$1 \leq N \leq 18; N = 20, 21, 28 \tag{4}$$

no hay ceros de ζ_N en el semiplano $\mathrm{Re}(s) > 1$, mientras que para los demás enteros positivos N, si existen infinitos ceros en esta región, los ceros especiales.
Sea

$$b_N = \sup\left\{\Re(s) : \zeta_N(s) = 0\right\} \tag{5}$$

Si $s = \sigma + it$ con $\sigma > 1$ y $\zeta_N(s) = 0$ entonces

$$1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{N^s} = 0$$

$$1 = \left|\frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{N^s}\right| \leq \frac{1}{2^\sigma} + \frac{1}{3^\sigma} + \cdots + \frac{1}{N^\sigma} < \zeta(\sigma) - 1$$

de donde $2 < \zeta(\sigma)$. Siendo ζ decreciente en el intervalo $\langle 1; +\infty\rangle$ y $\zeta(1{,}72865) = 2$, concluimos que $b_N \leq 1{,}73$

En 1983 H.L. Montgomery [7] demuestra que si $0 < c < \left(\frac{4}{\pi} - 1\right)$, $\exists N_0(c)/N > N_0(c)$, ζ_N tiene ceros en $\sigma > 1 + c\frac{\log(\log N)}{\log N}$, luego en el 2001 H.L. Montgomery y R. C. Vaughan [8] demuestran que existe $N_0/N > N_0$, ζ_N no tiene ceros en $\sigma \geq 1 + \left(\frac{4}{\pi} - 1\right)\frac{\log(\log N)}{\log N}$. Así

$$1 + c\frac{\log(\log N)}{\log N} < b_N \leq 1 + \left(\frac{4}{\pi} - 1\right)\frac{\log(\log N)}{\log N}$$

de donde

$$\lim_{N\to\infty} b_N = 1$$

Sea ahora

$$a_N = \text{ínf}\left\{\Re(s) : \zeta_N(s) = 0\right\} \tag{6}$$

En el 2009 M. Balazard y Oswaldo Velásquez [9] demuestran que $\lim_{N\to\infty}\frac{a_N}{N} = -\log(2)$, de donde

$$\lim_{N\to\infty} a_N = -\infty$$

de esta forma todos los ceros de ζ_N estan en la franja vertical $[a_N; b_N] \times \mathbb{R}$.
En el trabajo de G. Mora 2013 [10] se ve que el conjunto $\left\{\Re(s) : \zeta_N(s) = 0\right\}$ es eventualmente denso en $[a_N; b_N]$, es decir $\exists\, N_0/N \geq N_0$

$$\overline{\left\{\Re(s) : \zeta_N(s) = 0\right\}} = [a_N; b_N] \tag{7}$$

Si $n_N(T)$ denota la cantidad de ceros de $\zeta_N(s)$ en el rectángulo $[a_N; b_N] \times [0; T]$. Una estimativa que aparece en los trabajos de S.M. Gonek y A.H. Ledoan [11] es

$$\left|n_N(T) - \frac{T}{2\pi}\log(N)\right| < \frac{N}{2}$$

Por ejemplo para $T \in \left[\frac{2k\pi}{\log 2}, \frac{(2k+1)\pi}{\log 2}\right\rangle$, $n_2(T) = k$ ya que los ceros z_k de altura $\leq T$ de ζ_2 son

$$z_0 = \frac{\pi i}{\log(2)},\ z_1 = \frac{3\pi i}{\log(2)},\ \ldots,\ z_{k-1} = \frac{(2k-1)\pi i}{\log(2)}$$

Más aún, en mayo del 2014 G. Mora y J.M.Sepulcre [12] demostraron que, para cada entero $N \geq 2$ existe $T > 0$ tal que

$$n_N(T) = \left[\!\left[\frac{T\log(n)}{2\pi}\right]\!\right]$$

2.2. Proyección sobre un convexo

En el siguiente resultado H será un espacio vectorial con un producto escalar $\langle , \rangle$ que es completo con respecto a la norma inducida $\|.\|_{\langle,\rangle}$, es decir, H denotará un espacio de Hilbert.

Teorema 1 (Theorem 5.2. [13]). Sea $K \subset H$ un conjunto no vacío cerrado y convexo. Entonces para todo $w \in H$, existe un único $u \in K$ tal que

$$|w - u| = \min_{v\in K}|w - v| = \text{dist}(w, K). \tag{8}$$

Además, u está caracterizado por la propiedad

$$u \in K \text{ y } \langle w - u, v - u\rangle \leq 0 \ \ \forall v \in K \tag{9}$$

El elemento u dado en este teorema es llamado la proyección de w sobre K y es denotado por

$$u := \text{Proy}_K w$$

la desigualdad (9) dice que el producto escalar del vector $\overrightarrow{uw}$ con el vector $\overrightarrow{uv}$ ($v \in K$) es ≤ 0, es decir, el ángulo θ determinado por esos dos vectores es $\geq \pi/2$.

2.3. Error en la interpolación polinomial

Denotaremos por $\mathcal{P}_n$ al conjunto de todas las funciones polinomiales P con coeficientes reales o complejos de grado $\leq n$

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n$$

Teorema 2 (Theorem 2.1.1.1, [14]). Dados $n+1$ puntos arbitrarios

$$(x_i, f_i), \quad i = 0, 1, \ldots, n, \quad x_i \neq x_j \text{ para } i \neq j,$$

existe un único polinomio $P \in \mathcal{P}_n$ tal que

$$P(x_i) = f_i, \quad i = 0, 1, \ldots, n$$

En la página 39 de [14] se demuestra que tal polinomio viene dado por

$$P(x) = \sum_{i=0}^{n} f_i \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - x_k}{x_i - x_k} \tag{10}$$

llamado *polinomio de interpolación de Lagrange.* Para obtener el polinomio de interpolación completo se puede resolver el problema para subconjuntos del conjunto de puntos dado, y llegar al caso general recursivamente. En efecto, dado un conjunto de puntos (x_i, f_i), $i = 0, 1, \ldots, n$ denotaremos por $P_{i_0 i_1 \ldots i_k}$ al polinomio en $\mathcal{P}_k$ tal que:

$$\begin{aligned} &P_i(x) \equiv f_i, && i = 0, 1, 2, \ldots, n \\ &P_{i_0 i_1 \ldots i_k}(x_{i_j}) = f_{i_j}, && j = 0, 1, \ldots, k \ \ (k \geq 1) \end{aligned} \tag{11}$$

Proposición 1. Estos polinomios estan relacionados por la siguiente fórmula recursiva

$$P_{i_0 i_1 \ldots i_k}(x) \equiv \frac{(x - x_{i_0}) P_{i_1 i_2 \ldots i_k}(x) - (x - x_{i_k}) P_{i_0 i_1 \ldots i_{k-1}}(x)}{x_{i_k} - x_{i_0}} \tag{12}$$

Demostración. Denotando el lado derecho de (12) por $R(x)$, demostraremos que R tiene las propiedades características de $P_{i_0 i_1 \ldots i_k}$. El grado de R es claramente menor o igual que k. Por las definiciones de $P_{i_1 i_2 \ldots i_k}$ y $P_{i_0 i_1 \ldots i_{k-1}}$,

$$R(x_{i_0}) = P_{i_0 i_1 \ldots i_{k-1}}(x_{i_0}) = f_{i_0},$$

$$R(x_{i_k}) = P_{i_1 i_2 \ldots i_k}(x_{i_k}) = f_{i_k},$$

y

$$R(x_{i_j}) = \frac{(x_{i_j} - x_{i_0}) f_{i_j} - (x_{i_j} - x_{i_k}) f_{i_j}}{x_{i_k} - x_{i_0}} = f_{i_j}$$

para $j = 1, 2, \ldots, k-1$. Así $R = P_{i_0 i_1 \ldots i_k}$, en vista de la unicidad de la interpolación polinomial dado en el teorema 2. □

	k=0	1	2	3
x_0	$f_0 \equiv P_0(x)$			
		$P_{01}(x)$		
x_1	$f_1 \equiv P_1(x)$		$P_{012}(x)$	
		$P_{12}(x)$		$P_{0123}(x)$
x_2	$f_2 \equiv P_2(x)$		$P_{123}(x)$	
		$P_{23}(x)$		
x_3	$f_3 \equiv P_3(x)$			

Tabla 1. *El proceso recursivo*

Las dos primeras columnas del cuadro contienen las coordenadas de los puntos dados (x_i, f_i) (en este caso $0 \leq i \leq 3$). Las columnas subsiguientes se llenan mediante el cálculo de cada entrada de forma recursiva desde sus dos "vecinos" en la columna anterior, según la ecuación (12). Fijando ideas

$$P_{0123}(x) \equiv \frac{(x - x_0) P_{123}(x) - (x - x_3) P_{012}(x)}{x_3 - x_0}$$

Ejemplo 1. Para dos puntos (x_0, f_0) y (x_1, f_1)

$$P_0(x) \equiv f_0 \quad , P_1(x) \equiv f_1$$

y

$$\begin{aligned} P(x) \equiv P_{01}(x) &= \frac{(x - x_0) f_1 - (x - x_1) f_0}{x_1 - x_0} \\ &= f_0 + \left(\frac{f_1 - f_0}{x_1 - x_0} \right) (x - x_0) \end{aligned}$$

Ahora, dada una función f y algunos de sus valores

$$f_i := f(x_i), \quad i = 0, 1, \ldots, n$$

surge la siguiente interrogante ¿Qué tan buena será la interpolación polinomial $P(x) \equiv P_{01\ldots n}(x) \in \mathcal{P}_n$ para reproducir los valores de $f(x)$ cuando $x \neq x_i$?. En el siguiente resultado veremos que bajo ciertas condiciones, será posible acotar este error $f(x) - P(x)$

Teorema 3 (Theorem 2.1.4.1, [14]). Si la función f es derivable $(n+1)$ veces, entonces para todo $\overline{x}$ existe ξ en el menor intervalo $I[x_0, \ldots, x_n, \overline{x}]$ que contiene a $\overline{x}$ y todas las abscisas x_i, satisfaciendo

$$f(\overline{x}) - P_{01\ldots n}(\overline{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (\overline{x} - x_0)(\overline{x} - x_1) \cdots (\overline{x} - x_n) \tag{13}$$

En particular, cuando $n = 1$

$$f(\overline{x}) - P_{01}(\overline{x}) = \frac{f''(\xi)}{2} (\overline{x} - x_0)(\overline{x} - x_1) \tag{14}$$

3. Cuestiones metodológicas

Determinaremos ahora los ceros de una función holomórfa $f : \Omega \to \mathbb{C}$ en una región acotada, utilizando un método de bisección en el plano para aislar cada cero, y luego determinarlo usando el método de Newton. Por el

principio del argumento, si $\partial R \subset \Omega$ es una curva cerrada, simple, poligonal, parametrizada en sentido antihorario, el número de ceros de f dentro de R, $N(R)$ está dado por

$$N(R) = \frac{1}{2\pi}\Delta_{\partial R} \arg f(s).$$

y la función argumento $\arg : \mathbb{C}\backslash]-\infty, 0] \to \mathbb{C}$ por

$$\arg(z) = 2\arctan\left(\frac{\Im(z)}{|z| + \Re(z)}\right),\ z \in \mathbb{C}\backslash\,]-\infty, 0]$$

Si s_1, s_2 son dos números complejos verificando que $\left|\Delta_{[s_1,s_2]} \arg f(s)\right| < \pi$, podemos calcular la variación del argumento como

$$\Delta_{[s_1,s_2]} \arg f(s) = \arg\Big(\frac{f(s_2)}{f(s_1)}\Big).$$

Suponiendo que $\partial R = \bigcup_{i=1}^{n}[z_i, z_{i+1}]$ donde $z_{n+1} = z_1$, y para $i = 1, \ldots, n$, la variación del argumento es $\left|\Delta_{[z_i,z_{i+1}]} \arg f(s)\right| < \pi$, se tiene que

$$N(R) = \frac{1}{2\pi}\sum_{i=1}^{n}\Delta_{[z_i,z_{i+1}]} \arg f(s) = \frac{1}{2\pi}\sum_{i=1}^{n}\arg\Big(\frac{f(z_{i+1})}{f(z_i)}\Big).$$

Dada una curva cerrada, simple arbitraria $\mathcal{C} \subset \Omega$, nuestro primer paso será discretizarla "adecuadamente" como una poligonal R, es decir, encontrar una partición para la cual se cumpla que la variación del argumento sobre cada segmento de extremos z_i, z_{i+1} sea menor que π. Luego evaluaremos la variación del argumento a lo largo del segmento $[z_i, z_{i+1}]$ y cuando éste resulte ser mayor a π, elegiremos un punto interior en $[z_i, z_{i+1}]$, por ejemplo el punto medio y calcularemos nuevamente la variación del argumento para cada segmento. En caso falle, repetiremos este proceso hasta que la variación del argumento sobre cada segmento sea menor que π. El criterio dado en el teorema 2.1 [15] nos permitirá determinar adecuadamente los puntos z_i sobre la frontera ∂R, para calcular la variación del argumento de f sobre ∂R.
Aquí, $\langle x, y\rangle = \Re(x\bar{y})$ es el producto escalar usual o euclidiano sobre $\mathbb{C}$.

Teorema 4 (Ying and Katz). Sea $P(s)$ una función afín compleja cuya imagen no contiene al cero, $f(s)$ una función holomorfa sobre el intervalo $[s_1, s_2]$ y $R(s)$ dada por $R(s) = f(s) - P(s)$. Si

$$\min_{s\in[s_1,s_2]}\left|P(s)\right| > \max_{s\in[s_1,s_2]}\left|R(s)\right|, \tag{15}$$

entonces

$$\left|\Delta_{[s_1,s_2]} \arg f(s)\right| < \pi \ \text{ y } \ \Delta_{[s_1,s_2]} \arg f(s) = \arg\Big(\frac{f(s_2)}{f(s_1)}\Big).$$

Más precisamente, si $s_0 \in [s_1, s_2]$ es tal que $P(s_0) = \min_{s\in[s_1,s_2]}\left|P(s)\right| > 0$, para todo $s \in [s_1, s_2]$

$$\big\langle f(s), P(s_0)\big\rangle > 0.$$

Demostración. Como $P\big([s_1, s_2]\big)$ es un conjunto convexo que no contiene al cero, sea $P(s_0)$ la proyección del 0 sobre este conjunto, por una caracterización geométrica clásica de este punto, dada en el teorema 1, para todo $s \in [s_1, s_2]$

$$\big\langle P(s) - P(s_0), 0 - P(s_0)\big\rangle \le 0,$$

o bien $\big\langle P(s), P(s_0)\big\rangle \ge \left|P(s_0)\right|^2$.
Entonces

$$\begin{aligned}\big\langle f(s), P(s_0)\big\rangle &= \big\langle P(s) + R(s), P(s_0)\big\rangle \\ &= \big\langle P(s), P(s_0)\big\rangle + \big\langle R(s), P(s_0)\big\rangle \\ &\ge \left|P(s_0)\right|^2 - \left|R(s)\right|\cdot\left|P(s_0)\right| \\ &\ge \big(\min_{s\in[s_1,s_2]}\left|P(s)\right| - \max_{s\in[s_1,s_2]}\left|R(s)\right|\big)\left|P(s_0)\right| > 0\end{aligned}$$

□

También corroboramos, de modo distinto el resultado de Lema 2.1 de [15], como consecuencia de la fórmula del resto de Taylor en una variable.

Lema 1. Sea $f(s)$ una función holomorfa sobre el intervalo $[s_1, s_2]$, $P(s)$ y $R(s)$ dados por

$$P(s) = f(s_1) + \frac{f(s_2) - f(s_1)}{s_2 - s_1}(s - s_1)$$

y $R(s) = f(s) - P(s)$. Si $M(s_1, s_2)$ es tal que para todo $s \in [s_1, s_2]$,

$$\left|f''(s)\right| \le M(s_1, s_2),$$

entonces

$$\max_{s\in[s_1,s_2]}\left|R(s)\right| \le M(s_1, s_2)\frac{|s_1 - s_2|^2}{8}.$$

Demostración. En una variable esto podría ser una sencilla aplicación de la fórmula de interpolación de Lagrange, pero en variable compleja esto requiere de un retoque. En efecto, sea $u \in \mathbb{C}$ con $|u| = 1$ fijo arbitrario, y definimos $f_u : [0, 1] \to \mathbb{R}$ por

$$f_u(t) = \Big\langle f\big(z_1 + t(z_2 - z_1)\big), u\Big\rangle.$$

También definimos la interpolación (lineal) polinomial $P_u : [0, 1] \to \mathbb{R}$ tal que

$$P_u(0) = f_u(0) = \big\langle f(z_1), u\big\rangle, \quad P_u(1) = f_u(1) = \big\langle f(z_2), u\big\rangle,$$

así que

$$\begin{aligned}P_u(t) &= P_u(0) + t\big(P_u(1) - P_u(0)\big) \\ &= \Big\langle f(z_1) + t\big(f(z_2) - f(z_1)\big), u\Big\rangle \\ &= \Big\langle P\big(z_1 + t(z_2 - z_1)\big), u\Big\rangle.\end{aligned}$$

La fórmula del error en la interpolación polinomial (Theorem 2.1.4.1, p. 49, [14]), para $n = 1$, nos dá, la existencia, para cada $t \in]0, 1[$, de $\xi \in]0, 1[$ tal que

$$f_u(t) - P_u(t) = \frac{f_u''(\xi)}{2!}t(t - 1),$$

que se transforma en

$$\langle R(t),u\rangle = \frac{\left\langle f''\big(z_1+\xi(z_2-z_1)\big)(z_2-z_1)^2,u\right\rangle}{2}t(t-1).$$

Usando la cota $M(z_1,z_2)$ para $f''(s)$ en $[s_1,s_2]$, y teniendo en cuenta que $0\le t(1-t)\le 1/4$ para todo $t\in[0,1]$ obtenemos

$$\big|\langle R(t),u\rangle\big| \le \frac{1}{8}M(z_1,z_2)|z_2-z_1|^2.$$

Finalmente, para $t\in[0,1]$ es suficiente elegir u tal que $\langle R(t),u\rangle = \big|R(t)\big|$ y reescribir la desigualdad anterior □

Este lema nos ayudará reemplazar el cálculo del máximo de $\big|R(s)\big|$ por un cálculo más simple, a partir de alguna cota de $\big|f''(s)\big|$ en el intervalo $[s_1,s_2]$. De otro lado, por ser P una función afín lineal $P\big([z_1,z_2]\big)=[P(z_1),P(z_2)]$, y para cualquier $t\in[0,1]$

$$P(z_1+t(z_2-z_1)) = P(z_1)+t(P(z_2)-P(z_1))$$

En el siguiente lema, usando los mismos argumentos del Lema 2.2 de [15], demostraremos que

$$\min_{\in[z_1,z_2]}|P(z)| \in \left\{\frac{\Im\big(P(z_1)\overline{P(z_2)}\big)}{|P(z_1)-P(z_2)|},\,|P(z_1)|,\,|P(z_2)|\right\} \quad (16)$$

Lema 2. Sea $a,b\in\mathbb{C}$ con $a\ne b$ y $m=\min\limits_{t\in[0,1]}|a+t(b-a)|$

1. Si $\Re(a\bar b)\ge|a|^2$ y $|b|>|a|$, entonces $m=|a|$.
2. Si $\Re(a\bar b)\ge|b|^2$ y $|a|>|b|$, entonces $m=|b|$.
3. En cualquier otro caso,
$$m=\frac{\big|\Im(a\bar b)\big|}{|b-a|}$$

Demostración. Sea $t^*\in\mathbb{R}$ tal que

$$\min_{t\in\mathbb{R}}\big|a+t(b-a)\big| = a+t^*(b-a),$$

entonces

$$0=\langle b-a,a+t^*(b-a)\rangle = \langle b,a\rangle - |a|^2+t^*|b-a|^2$$
$$= \Re(a\bar b)-|a|^2+t^*|b-a|^2$$

de donde

$$t^*=\frac{|a|^2-\Re(a\bar b)}{|b-a|^2}$$

Caso 1. $\Re(a\bar b)\ge \min\{|a|^2,|b|^2\}$

- $\Re(a\bar b)\ge|a|^2$ y $|b|\ge|a|$, entonces $t^*\le 0$, de donde $t^*=0$ y $\min\limits_{t\in[0,1]}\big|a+t(b-a)\big|=|a|$.
- Si $\Re(a\bar b)\ge|b|^2$ y $|a|\ge|b|$, entonces $t^*=\dfrac{|a|^2-\Re(a\bar b)}{|b-a|^2}=\dfrac{|a|^2-\langle a,b\rangle}{|b-a|^2}\ge 1$, puesto que

$$\langle a,b\rangle\ge|b|^2 \Leftrightarrow |a|^2+\langle a,b\rangle\ge|a|^2+|b|^2$$
$$\Leftrightarrow |a|^2-\langle a,b\rangle\ge|a|^2+|b|^2-2\langle a,b\rangle=|b-a|^2$$

de donde $t^*=1$ y $\min\limits_{t\in[0,1]}\big|a+t(b-a)\big|=|b|$.

Caso 2. $\Re(a\bar b)<\min\{|a|^2,|b|^2\}$, por tanto $t^*\in\langle 0,1\rangle$ y

$$|a+t^*(b-a)||\overline{b-a}| = \big|a\bar b-|a|^2+t^*|b-a|^2\big|$$
$$= |a\bar b-\Re(a\bar b)| = |i\Im(a\bar b)|$$

Así

$$\min_{t\in[0,1]}\big|a+t(b-a)\big| = \frac{|\Im(a\bar b)|}{|b-a|}$$

□

Con estas herramientas a la mano, nuestro criterio de discretización para una curva cerrada simple $\mathcal{C}$ será dividir este contorno, en un contorno poligonal de segmentos $[s_1',s_2']$ tales que

$$\min_{s\in[s_1',s_2']}|P(s)| > M(s_1',s_2')\frac{|s_1'-s_2'|^2}{8}.$$

Esto, junto con el Lema 1, da la condición del Teorema 4 para $f(s)$ sobre el intervalo $[s_1',s_2']$.

Necesitamos ahora calcular una cota superior para la segunda derivada de cada uno de las funciones que aquí investigamos, esto es, funciones de la forma

$$f(z)=\sum_{k=1}^{n}a_ke^{\lambda_k z}\;;\quad z\in[z_1,z_2],\quad \lambda_k<0$$

Como $f''(z)=\sum\limits_{k=1}^{n}a_k\lambda_k^2e^{\lambda_k z}$, entonces

$$|f''(z)|\le\sum_{k=1}^{n}|a_k|\lambda_k^2e^{\Re(\lambda_k z)} \text{ para } z\in[z_1,z_2],$$

esto es $z=(1-t)z_1+tz_2$ con $0\le t\le 1$, de donde

$$\Re(\lambda_k z)\le \max_{z\in[z_1,z_2]}\Re(\lambda_k z)$$
$$= \max_{t\in[0,1]}\big\{(1-t)\lambda_k\Re(z_1)+t\lambda_k\Re(z_2)\big\}$$
$$\le \lambda_k\min\{\Re(z_1),\Re(z_2)\}$$

entonces

$$|f''(z)|\le\sum_{k=1}^{n}|a_k|\lambda_k^2e^{\lambda_k\min\{\Re(z_1),\Re(z_2)\}},\quad z\in[z_1,z_2]$$

es decir, $|f''(z)|\le M(z_1,z_2)$ para

$$M(z_1,z_2)=\sum_{k=1}^{n}|a_k|\lambda_k^2e^{\lambda_k\min\{\Re(z_1),\Re(z_2)\}} \quad (17)$$

En particular, para

$$\zeta_n(z)=\sum_{k=1}^{n}\frac{1}{k^z}=\sum_{k=1}^{n}e^{-\log(k)z}$$

$$M(z_1,z_2)=\sum_{k=2}^{n}[\log(k)]^2e^{-\log(k)\min\{\Re(z_1),\Re(z_2)\}} \quad (18)$$

4. Resultados

Para cada altura T, de acuerdo a las técnicas utilizadas en el procesamiento de los datos, denotando con $n_N(T)$ al número de ceros de ζ_N en el rectángulo $[a_N(T), b_N(T)] \times [0, T]$, es decir

$$n_N(T) = \left|\left\{s \in \mathbb{C} \,/\, \zeta_N(s) = 0,\ 0 < \Im(s) < T\right\}\right|$$

y los términos $a_N(T), b_N(T)$ por

$$a_N(T) = \text{ínf}\left\{\, \Re(s) \ / \ \zeta_N(s) = 0,\ 0 < \Im(s) < T \,\right\} \quad (19)$$

y

$$b_N(T) = \sup\left\{\, \Re(s) \ / \ \zeta_N(s) = 0,\ 0 < \Im(s) < T \,\right\} \quad (20)$$

El resultado presentado en la siguiente figura ha sido generado usando un código escrito en Python 3, mientras que la gráfica de la nube de ceros ha sido generado usando gnuplot en formato eps. Para poder mantener una buena precisión en los valores procesados, se decidió usar la librería Mpmath [16], la cual permite realizar cálculos con números reales y complejos manteniendo una alta precisión (10 dígitos o 1000 digitos), por ejemplo aquí se usó una precisión de 100 dígitos significativos en el entorno de Mpmath.

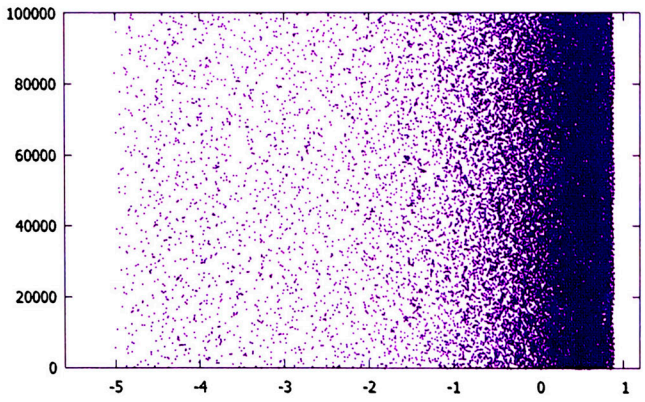


Figura 2. *Nube de ceros de* $\zeta_9(s)$

En el siguiente cuadro se muestra los resultados obtenidos en esta investigación para los distintos valores de N y con cierta altura prefijada.

N	T	$n_N(T)$	$\frac{T}{2\pi}\log(N)$	$a_N(T)$	$b_N(T)$
3	10^6	174850	174849.576	-0.999999	0.787884
4	10^6	220636	220635.600	-1.214285	0.626288
5	10^6	256150	256149.999	-2.425976	0.890899
6	10^6	285168	285167.376	-2.886476	0.841088
7	10^5	30970	30970.122	-3.802603	0.975899
8	10^5	33095	33095.340	-4.378860	0.919274
9	10^5	34970	34969.915	-5.009246	0.893349

Tabla 2. *Recopilación de los datos procesados*

Por ejemplo, para la función $\zeta_9(s)$ vemos que la región de estudio es $[-5{,}009; 0{,}893] \times [0; 100000]$, y en ella hay 34 970 ceros, que están bien aproximados por $\frac{10^5}{2\pi}\log(9) = 34969{,}915$.

5. Conclusiones

- El clásico problema de Rieman es aún una gran fuente de inspiración para seguir generando mejoras, y quizá nuevos resultados.
- Sobre los ceros de las funciones enteras ζ_N, tenemos el legado que nos dejan: Turan, Spira, Monach, Montgomery, Vaughan, Balazard, Oswaldo Velásquez, Mora, Gonek, Ledoan, Sepulcre, Platt, Trudgian, etc.
- Este trabajo está basada en los resultados de X. Ying y I.N. Katz, cuyas demostraciones fueros aquí mejoradas usando las herramientas del análisis convexo y análisis numérico.
- Para la implementación numérica se ha usado Python 3, logrando generar una data efectiva de ceros de ζ_N, cuyo análisis corrobora resultados analíticos de nuestros predecesores, y seguirán siendo explorados con la confianza de descifrar la matemática que allí existe.

Agradecimiento

Nuestro agradecimiento al Instituto de Investigación de la Facultad de Ciencias de la UNI, por su apoyo con esta publicación.

1. Platt, David J and Trudgian, Timothy S, LMS Journal of Computation and Mathematics, 19, 1, 37–41, 2016, London Mathematical Society.
2. Elchin Hasanalizade and Quanli Shen and Peng-Jie Wong, Journal of Number Theory,2021, https://doi.org/10.1016/j.jnt.2021.06.032, https://www.sciencedirect.com/science/article/pii/S0022314X2100233X.
3. Turán, Paul, 24, 17, 1948, I kommission hos Munksgaard.
4. Spira, Robert, Mathematics of Computation, 20, 96, 542–550, 1966, JSTOR.
5. Numerical investigation of several problems in number theory, Univ. of Michigan Ph. D Dissertation, William Reynolds, Ann Arbor, 1980.
6. Platt, David J and Trudgian, Timothy S, LMS Journal of Computation and Mathematics, 19, 1, 37–41, 2016, London Mathematical Society.
7. Montgomery, Hugh L, Studies in pure mathematics, 497–506, 1983, Springer.
8. Montgomery, Hugh L and Vaughan, Robert C, Periodica Mathematica Hungarica, 43, 1-2, 199–214, 2002, Springer.

9. Balazard, Michel and Castanón, Oswaldo Velásquez, Comptes Rendus Mathematique, 347, 7-8, 343–346, 2009, Elsevier

10. Mora, Gaspar, arXiv preprint arXiv:1311.5171, 2013

11. Gonek, Steven M and Ledoan, Andrew H, International Mathematics Research Notices, 2010, 10, 1775–1791, 2010, OUP.

12. Mora, Gaspar and Sepulcre, Juan Matias, Annali di Matematica Pura ed Applicata (1923-), 194, 5, 1499–1504, 2015, Springer.

13. Functional analysis, Sobolev spaces and partial differential equations, Brezis, Haim, 2010, Springer Science & Business Media.

14. Introduction to numerical analysis, Stoer, Josef and Bulirsch, Roland, 12, 2013, Springer Science & Business Media.

15. Ying, Xingren and Katz, I Norman, Numerische Mathematik, 53, 1-2, 143–163, 1988, Springer.

16. Fredrik Johansson and others, mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18), `http://mpmath.org/`, December, 2013.

Un problema de Optimización y las condiciones de optimalidad de Karush Khun Tucker

Johnny M. Valverde Montoro[†]
Escuela Profesional de Matemática. Facultad de Ciencias.
Universidad Nacional de Ingeniería;
[†]*jvalverde@uni.edu.pe*

Recibido el 19 de mayo del 2020; aceptado el 21 de agosto del 2020

El presente trabajo muestra las condiciones de Karush-Khun-Tucker en problemas de optimización multiobjetivo con funciones objetivo de valor intervalo considerando relaciones de orden parcial sobre la familia de todos los intervalos cerrados en $\mathbb{R}$. Se emplean elementos de la aritmética de intervalos y la diferencia generalizada de Hukuhara.

Palabras clave: KKT, función multivalor intervalo.

This work shows the conditions of Karush-Khun-Tucker in multiobjective optimization problems with objective functions of interval value considering partial order relationships on the family of all closed intervals in $\mathbb{R}$. In the study carried out the interval arithmetic is used and generalized difference of Hukuhara.

Keywords: KKT, funciones multivalor intervalo.

1. Introducción

La imprecisión es algo inevitable en situaciones inesperadas, en consecuencia, considerar la incertidumbre dentro de los problemas de optimización definen una línea de investigación. Los problemas de programación lineal que presentan inexactitud estan muy relacionadas a los problemas de optimización que emplean valores intervalo.

Ishibuchi y Tanaka [5] estudiaron los problemas de programación multiobjetivo con funciones de valor intervalo y propusieron una relación de orden entre dos intervalos cerrados. Los problemas de optimización matemática con funciones objetivo de valor intervalo son estudiadas por Wu [9], empleando la diferencia de Hukuhara y la Aritmética de intervalos se define la Derivada de Hukuhara, conocida como H-derivada, y presenta dos relaciones parciales sobre las cuales plantea las condiciones de optimalidad de Karush Khun Tucker (KKT) en un problema de optimización empleando funciones de valor intervalo. En base a este trabajo se extiende este estudio a las funciones multiobjetivo de valor intervalo y los presenta en el articulo de Wu [10]. Asimismo, en el trabajo de Hosseinzade [4] retoma los estudios de Wu, incidiendo sobre las relaciones de orden parcial planteadas por este investigador.

La diferencia de Hukuhara es extendida a la que se conoce como la diferencia generalizada de Hukuhara y de este modo la H-Derivada da paso a la que se conoce como la Derivada Generalizada de Hukuhara o simplemente gH-derivada presentado por Stefanini [7]. Chalco [3] presenta nuevas relaciones de orden parcial para replantear las condiciones en los problemas de optimización de funciones de valor intervalo aplicando la gH-derivada, tomando como referencia las relaciones de orden parcial trabajadas por Wu [9]. Se desarrolla una extensión a conjuntos difusos por Stefanini [8] empleando la diferencia generalizada de Hukuhara. En este trabajo Se trata de estudiar la existencia de una solución óptima de un problema de optimización para el caso de funciones multivalor de valor intervalo empleando una reformulación de las condiciones de Karush Khun Tucker considerando la derivada generalizada de Hukuhara.

2. Preliminares

Sea $\mathscr{I}$ la clase de todos los intervalos cerrados y acotados en $\mathbb{R}$. Sobre $\mathscr{I}$ se operan sus elementos en base a la aritmética de intervalos (Moore [6]).

2.1. Diferencia de Hukuhara

Sean $A = [a^I, a^S]$ y $B = [b^I, b^S]$ dos elementos de $\mathscr{I}$. Si existe un intervalo cerrado $C = [c^I, c^S]$ de modo que $A = B + C$, entonces C es llamada la diferencia de Hukuhara. Desde que $A = B + C$, no es difícil ver que $a^I = b^I + c^I$ y $a^S = b^S + c^S$, esto es, $c^I = a^I - b^I$ y $c^I = a^I - b^I$. Por lo tanto, este intervalo cerrado C existe si $a^I - b^I \leq a^S - b^S$, esto el ancho de B no supere al ancho de A. En este caso, $C = [a^I - b^I, a^S - b^S]$ y escribimos $C = A \ominus B$. En consecuencia, cuando decimos que la diferencia de Hukuhara $C = A \ominus B$ existe, implícitamente significa que $a^I - b^I \leq a^S - b^S$.

El ancho de un intervalo cerrado $A = [a^I, a^S]$ en $\mathbb{R}$, denotado por $w(A)$, es la cantidad $w(A) = a^S - a^I$. Se puede notar que no siempre existe esta diferencia.

Ejemplo 1. *sean* $A = [1, 2]$ *y* $B = [2, 5]$. *Como el ancho de* A, $w(A) = 1$, *es menor que el ancho de* B, $w(B) = 3$ *entonces no existe* $C = A \ominus B$

De esto se puede notar que no siempre esta definida la diferencia de Hukuhara.

2.2. Diferencia Generalizada de Hukuhara

Como la diferencia de Hukuhara es muy restrictiva, esto conlleva a tener que replantearla y proponer la denominada Diferencia Generalizada de Hukuhara.

Definición 2.1. *(Stefanini y Bede [7]) La Diferencia Generalizada de Hukuhara, denominada gH-diferencia, entre dos intervalos cerrados $C = [c^I, c^S]$ y $D = [d^I, d^S]$, elementos de $\mathscr{I}$, se define*

$$C \ominus_g D = E \Leftrightarrow \begin{cases} C = D + E, & si\ w(C) \geq w(D) \\ D = C + (-1)E, & si\ w(C) < w(D) \end{cases} \tag{1}$$

donde $w(C)$ y $w(D)$ son los anchos de los intervalos cerrados C y D respectivamente.

Esta nueva diferencia tiene propiedades interesantes, se puede observar que para cualquier $C \in \mathscr{I}$ se cumple $C \ominus_g C = \{0\} = [0, 0]$, donde los conjuntos unitarios $\{0\}$ en $\mathbb{R}$ se consideraran como intervalos cerrados degenerados, denotados por $[a, a]$. Asimismo, la gH-diferencia siempre existe entre dos intervalos cerrados cualesquiera en $\mathbb{R}$, esto es, para cualquier par de elementos de $\mathscr{I}$ se obtiene la gH-diferencia.

Proposición 2.1. *(Stefanini y Bede [7]) Sean los intervalos cerrados $C = [c^I, c^S]$ y $D = [d^I, d^S]$, elementos cualesquiera de $\mathscr{I}$, se tiene que*

$$\begin{aligned} C \ominus_g D = &[Min\left\{c^I - d^I, c^S - d^S\right\}, \\ &Max\left\{c^I - d^I, c^S - d^S\right\}] \end{aligned} \tag{2}$$

Ejemplo 2. *sean $A = [-1, 4]$ y $B = [1, 3]$,*
$C = A \ominus_g B = [Min\{-1-1, 4-3\}, Max\{-1-1, 4-3\}]$
$= [Min\{-2, 1\}, Max\{-2, 1\}] = [-2, 1]$,
donde $w(B) = 2 < w(A) = 5$

Ejemplo 3. *sean $A = [1, 3]$ y $B = [2, 5]$,*
$C = A \ominus_g B = [Min\{1-2, 3-5\}, Max\{1-2, 3-5\}]$
$= [Min\{-1, -2\}, Max\{-1, -2\}] = [-2, -1]$, *donde* $w(A) = 2 < w(B) = 3$

Proposición 2.2. *sean $A, B, C \in \mathscr{I}$ cualesquiera. Se cumple*

(i) $k(A \ominus_g B) = kA \ominus_g kB$, $k \in \mathbb{R}$

(ii) $A \ominus_g B = [0, 0]$ *si y solo si* $A = B$

(iii) $(A + B) \ominus_g (A + C) = B \ominus_g C$

Demostración. (i) y (ii) se obtienen de la definición 2.1. (iii) sea $D \in \mathscr{I}$ de modo que $(A+B) \ominus_g (A+C) = D$. Por la definición de la gH-diferencia se cumple que $A + B = A + C + D$ o $A + C = A + B + (-1)D$. De esto, se tiene que $B = C + D$ o $C = B + (-1)D$ y por definición de la gH-diferencia se tiene que $B \ominus_g C$ □

2.3. Diferenciabilidad de funciones de valor intervalo

Considerando los elementos del cálculo clásico se tiene que para un abierto $X \subset \mathbb{R}$, $x_0 \in X$ y una función de valor intervalo $F : X \to \mathscr{I}$, con $F(x) = [F^I(x), F^S(x)]$, donde F^I y F^S son funciones reales sobre X, se dice que F es diferenciable en x_0 siempre que las funciones reales F^I y F^S sean diferenciables en x_0 (en el sentido usual). En base a la diferencia de Hukuhara se plantea el siguiente tipo de diferenciabilidad (Wu [9]):

Definición 2.2. *Sea X un abierto en $\mathbb{R}$. Una función de valor intervalo $F : X \to \mathscr{I}$ es denominada H-*diferenciable *(o* fuertemente diferenciable*) en $x_0 \in \mathbb{R}$ si existe un intervalo cerrado $A(x_0) \in \mathscr{I}$ (el cual depende de x_0) de modo que los límites*

$$\lim_{h \to 0^+} \frac{F(x_0 + h) \ominus F(x_0)}{h} \quad y \quad \lim_{h \to 0^+} \frac{F(x_0) \ominus F(x_0 - h)}{h}$$

*ambos existen y son iguales a $A(x_0)$. En este caso, $A(x_0)$ es llamada la H-*derivada *de F en x_0.*

Ejemplo 4. *Sea $F : \mathbb{R} \to [(x-1)^2+1, x^2+2]$ una función de valor intervalo definida sobre $\mathbb{R}$. Se puede ver que F es H-diferenciable en cualquier $x_0 \in \mathbb{R}$, con H-derivada $[2x_0 - 2, 2x_0]$.*

Se debe notar que cuando F es H-diferenciable en x_0, de manera implícita se tiene que $F(x_0 + h) \ominus F(x_0)$ y $F(x_0) \ominus F(x_0 - h)$ existen para todo $h > 0$.
Pero, la Derivada de Hukuhara es muy restrictiva debido a que la Diferencia de Hukuhara no siempre existe, ya que esta limitada por su propia definición. Por tal motivo, se defne la Derivada Generalizada de Hukuhara.

Definición 2.3. *Sea X un abierto en $\mathbb{R}$, $t_0 \in X$. La Derivada generalizada de Hukuhara, denominada gH-derivada, de una función de valor intervalo $F : X \to \mathscr{I}$ en t_0, está definida como*

$$F'(t_0) = \lim_{h \to 0} \frac{F(t_0 + h) \ominus_g F(t_0)}{h}. \tag{3}$$

Si existe $F'(t_0) \in \mathscr{I}$ satisfaciendo (3), entonces se dice que F es $gH-$diferenciable en t_0. Se dice que la función de valor intervalo $F : X \to \mathscr{I}$ es $gH-$diferenciable en X si F es $gH-$diferenciable en cada $t_0 \in X$. El siguiente resultado, presentado por Chalco [3], expresa la $gH-$derivada en términos de los extremos de la imagen (intervalo cerrado) de la función valor intervalo.

Teorema 2.3. *Sea X un abierto en $\mathbb{R}$, $t_0 \in X$ y $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$. Si F^I y F^S son funciones diferenciables en $t_0 \in X$, entonces F es $gH-$diferenciable en t_0 y*

$$\begin{aligned} F'(t_0) = &\left[\text{mín}\left\{(F^I)'(t_0), (F^S)'(t_0)\right\},\right. \\ &\left.\text{máx}\left\{(F^I)'(t_0), (F^S)'(t_0)\right\}\right]. \end{aligned} \tag{4}$$

Se debe notar que el recíproco del Teorema 2.3 no es cierto, es decir, la $gH-$diferenciabilidad de F no implica la diferenciabilidad de F^I y F^S (en el sentido usual). Por

ejemplo, si se considera la función F de valor intervalo, dada por $F(t) = [-|kt|, |kt|], k \neq 0$, se tiene que F es $gH-$diferenciable en $t_0 = 0$ y $F'(0) = [-k, k]$. Sin embargo F^I y F^S no son funciones diferenciables en $t_0 = 0$. En general, se tiene el siguiente resultado:

Teorema 2.4. *(Chalco [3]) Sea X un abierto en $\mathbb{R}$, $t_0 \in X$ y $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$. Se tiene que, F es $gH-$diferenciable en $t_0 \in X$ cuando y sólo cuando uno de los siguientes casos es válido*

(a) F^I y F^S son diferenciables en t_0;

(b) las derivadas laterales $(F^I)'_-(t_0), (F^I)'_+(t_0)$, $(F^S)'_-(t_0)$ y $(F^S)'_+(t_0)$ existen y satisfacen $(F^I)'_-(t_0) = (F^S)'_+(t_0)$ y $(F^I)'_+(t_0) = (F^S)'_-(t_0)$.

A partir de este teorema se deduce la siguiente proposición:

Proposición 2.5. *Sea X un abierto en $\mathbb{R}$, $t_0 \in X$ y $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$. Si F es $gH-$diferenciable en t_0, entonces $(F^I + F^S)$ es una función diferenciable en t_0.*

Se extenderá el estudio de las funciones valor intervalo sobre $\mathbb{R}^n$, esto es, $F(\mathbf{x}) = F(x_1, \ldots, x_n)$ es un intervalo cerrado en $\mathbb{R}$, para cada $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. En consecuencia, también se tienen las correspondientes funciones reales $F^I(\mathbf{x}) = F^I(x_1, \ldots, x_n)$ y $F^S(\mathbf{x}) = F^S(x_1, \ldots, x_n)$ definidas sobre $\mathbb{R}^n$, de modo que $F(\mathbf{x}) = [F^I(\mathbf{x}), F^S(\mathbf{x})] \in \mathscr{I}$.

Proposición 2.6. *Sea X un abierto en $\mathbb{R}^n$, $x_0 \in X$ y $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$. Se cumple que F es continua en x_0 si y solo si F^I y F^S son continuas en x_0.*

Recordando del Cálculo la siguiente afirmación:

Proposición 2.7. *Sea f una función real definida sobre $\mathbb{R}^n$. Si asumimos que una de las derivadas parciales $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ existe en $\mathbf{x}_0$ y que las $n-1$ derivadas parciales restantes existen en alguna vecindad de $\mathbf{x}_0$ y son continuas en $\mathbf{x}_0$, entonces f es diferenciable en $\mathbf{x}_0$.*

Se plantea la siguiente definición:

Definición 2.4. *Sea F una función de valor intervalo definida en $X \subset \mathbb{R}^n$ y $x_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$ un elemento de X fijo. Se dice que F es continuamente $gH-$diferenciable en x_0 si todas las $gH-$derivadas parciales $(\frac{\partial F}{\partial x_1})_g(x_0), \ldots, (\frac{\partial F}{\partial x_n})_g(x_0)$ existen en alguna vecindad de x_0 y son continuas en x_0 (en el sentido de función valor intervalo).*

Definición 2.5. *Sea X un abierto en $\mathbb{R}^n$, $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$, $x_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$ un elemento de X fijo y la función de valor intervalo h_i definida por $h_i(x_i) = F(x_1^{(0)}, \ldots, x_{i-1}^{(0)}, x_i, x_{i+1}^{(0)}, \ldots, x_n^{(0)})$ Si h_i es $gH-$diferenciable en $x_i^{(0)}$, entonces se dice que F tiene la i-ésima $gH-$derivada parcial en x_0 (denotado por $(\frac{\partial F}{\partial x_i})_g(x_0)$, donde $(\frac{\partial F}{\partial x_i})_g(x_0) = (h_i)'(x_i^{(0)})$.*

Proposición 2.8. *Sea F una función de valor intervalo definida en $X \subset \mathbb{R}^n$, donde $F(x) = [F^I(x), F^S(x)]$, para $x \in X$. Si F es continuamente $gH-$diferenciable en x_0, entonces $(F^I + F^S)$ es continuamente diferenciable en x_0.*

Demostración. Como $F(x) = [F^I(x), F^S(x)]$, $x \in \mathbb{R}^n$. En el caso de que si $(\frac{\partial F}{\partial x_i})_g(x_0)$ existe, entonces, de la Proposición 2.5, la derivada parcial $\frac{\partial}{\partial x_i}(F^I + F^S)(x_0)$ existe. Por otro lado, de la Proposición 2.5 y 2.6, como $(\frac{\partial F}{\partial x_i})_g(x_0)$ es continua, entonces $\frac{\partial}{\partial x_i}(F^I + F^S)$ es continua en x_0. En consecuencia, si F es continuamente $gH-$diferenciable en x_0, entonces la función de valor real $(F^I + F^S)$ es continuamente diferenciable en x_0. □

A continuación se presenta una definición que ha sido tomada de Chalco [3], la cual es una reformulación de la presentada por Wu [9].

Definición 2.6. *Se dice que la función de valor intervalo $F : X \to \mathscr{I}$ es (débilmente) continuamente diferenciable en $x_0 \in X$ si las funciones de valor real F^I y F^S son continuamente diferenciables en x_0 , esto es, todas las derivadas parciales de F^I y F^S existen en algunas vecindades de x_0 y son continuas en x_0 (en el sentido usual).*

Definición 2.7. *Sea X un abierto en $\mathbb{R}^n$, $t_0 = \left(t_1^{(0)}, \ldots, t_n^{(0)}\right)$ un elemento de X fijo y $F : X \to \mathscr{I}$ una función de valor intervalo de modo que $F(t) = [F^I(t), F^S(t)]$. El gradiente gH de F en t_0, denotado por $\nabla_g F(\mathbf{t}_0)$, está definido por*

$$\nabla_g F(\mathbf{t}_0) = \left(\left(\frac{\partial F}{\partial t_1}\right)_g (\mathbf{t}_0), \ldots, \left(\frac{\partial F}{\partial t_n}\right)_g (\mathbf{t}_0) \right)$$

donde $(\frac{\partial F}{\partial t_j})_g(\mathbf{t}_0)$ es la j-ésima gH-derivada parcial de f en $\mathbf{t}_0$ $(j = 1, 2, \ldots, n)$, como fue definido en la Definición 2.5. Se puede observar que si las funciones extremas F^I y F^S son funciones diferenciables, entonces F es gH-diferenciable y en este caso

$$\left(\frac{\partial F}{\partial x_j}\right)_g (\mathbf{x}_0) = [A, B]$$

es un intervalo cerrado, con $j = 1, 2, \ldots, n$ donde:

$$A = min\left\{ \frac{\partial F^I}{\partial x_j}(\mathbf{x}_0), \frac{\partial F^S}{\partial x_j}(\mathbf{x}_0) \right\}$$

$$B = max\left\{ \frac{\partial F^I}{\partial x_j}(\mathbf{x}_0), \frac{\partial F^S}{\partial x_j}(\mathbf{x}_0) \right\}$$

Ejemplo 5. *Considere la función F de valor intervalo definida por*

$$F(x) = F(x_1, x_2) = \left[x_1 + {x_2}^2, {x_1}^2 + {x_2}^2 + 3\right].$$

Entonces tenemos

$$\left(\frac{\partial F}{\partial x_1}\right)_g (x_1, x_2) = [min\{1,2x_1\}, max\{1, 2x_1\}]$$

y

$$\left(\frac{\partial F}{\partial x_2}\right)_g (x_1, x_2) = [min\{2x_2, 2x_2\}, max\{2x_2, 2x_2\}] = [2x_2, 2x_2]$$

Así, el gradiente gH de F está dada por

$$\nabla_g F(x_1, x_2) = ([min\{1, 2x_1\}, max\{1, 2x_1\}], [2x_2, 2x_2]).$$

Ahora, se considerarán las funciones multivalor intervalo F, las cuales están definidas sobre el abierto $X \subset \mathbb{R}^n$, con $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$, donde cada una de las $F_i : X \to \mathscr{I}$ son funciones de valor intervalo, para $i = 1, 2, \ldots, q$, esto es, $F_i(t) = [F_i^I(t), F_i^S(t)]$, $t \in X$, para $i = 1, 2, \ldots, q$.

Ejemplo 6. *sea F una función multivalor intervalo definida sobre $\mathbb{R}^2$ dada por $F(x_1, x_2) = ([x_1^2 + 2x_1x_2, x_1 + x_2^2 + 3], [x_1^2 + 3, 3x_1x_2])$*
donde: $F_1(x_1, x_2) = [x_1^2 + 2x_1x_2, x_1 + x_2^2 + 3]$
$F_2(x_1, x_2) = [x_1^2 + 3, 3x_1x_2]$
F_1 y F_2 son funciones valor intervalo en $\mathbb{R}^2$

Definición 2.8. *Sea F una función multivalor definida sobre el abierto $X \subset \mathbb{R}^n$, con $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$. Se dice que F es*

a) (debilmente) continuamente diferenciable en $x_0 \in X$ si F_i es (debilmente) continuamente diferenciable en x_0, para cada $i = 1, 2, \ldots, q$

b) continuamente gH-diferenciable en $x_0 \in X$ si F_i es continuamente gH-diferenciable en x_0, para cada $i = 1, 2, \ldots, q$

Proposición 2.9. *Sea F una función multivalor intervalo definida sobre el abierto $X \subset \mathbb{R}^n$, con $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$, donde $F_i(t) = [F_i^I(t), F_i^S(t)]$, $t \in X$, para $i = 1, 2, \ldots, q$. Se cumple que*

a) si F_i^I, F_i^S son diferenciables en x_0, para cada $i = 1, 2, \ldots, q$, entonces F es (debilmente) continuamente diferenciable en $x_0 \in X$.

b) si F es continuamente gH-diferenciable en $x_0 \in X$, entonces $(F_i^I + F_i^S)$ es continuamente diferenciable en x_0, para cada $i = 1, 2, \ldots, q$.

Demostración. (i) se obtiene de las definiciones 2.6 y 2.8 (ii) es consecuencia de la definición 2.8 y de la proposición 2.8. □

3. Formulación del problema de investigación

Las condiciones de optimalidad de Karush Khun Tucker se tienen que replantear para las funciones de valor intervalo. Para lograr esto se deben de proponer nuevas relaciones de orden parcial entre los miembros de la familia de intervalos cerrados $\mathscr{I}$, las cuales serán extendidas a los vectores intervalo, cuyos componentes serán elementos de $\mathscr{I}$.
La presente investigación se enmarca en un estudio que involucra la diferenciabilidad aplicada a funciones de valor intervalo teniendo que emplear un nuevo tipo de derivada: la derivada H (de Hukuhara). Este tipo de derivada se basa sobre la diferencia de Hukuhara, la cual es muy restrictiva, motivo por el cual se debe extender a la diferencia generalizada de Hukuhara lo que permitirá definir la denominada derivada generalizada de Hukuhara conocida como la gH-Derivada, Stefanini [7]. En los problemas de optimización lo que se busca es minimizar o maximizar una función objetivo y en el presente estudio el problema de investigación lo formulamos por medio de la siguiente pregunta: En los problemas de optimización lo que se busca es minimizar o maximizar una función objetivo y en el presente estudio el problema de investigación lo formulamos por medio de la siguiente pregunta:

¿Existe solución para el siguiente programa matemático

$$\begin{array}{lll} \text{(P)} & \text{mín} & F(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_q(\mathbf{x})) \\ & \text{sujeto a} & g_i(\mathbf{x}) \leqslant 0, \quad i = 1, \ldots, m \end{array}$$

de modo que satisfaga las condiciones de optimalidad de Karush Khun Tucker donde F es una función multivalor intervalo (sus componentes son funciones de valor intervalo)?

4. Optimización de funciones de valor intervalo

Se tiene los siguientes programas matemáticos:

$$\begin{array}{lll} \textbf{(IP1)} & \text{mín} & f(x) = [f^I(x), f^S(x)] \\ & \text{sujeto a} & x = (x_1, \ldots, x_n) \in X \subseteq \mathbb{R}^n. \end{array}$$

$$\begin{array}{lll} \textbf{(IP2)} & \text{mín} & f(x) = [f^I(x), f^S(x)] \\ & \text{sujeto a} & g_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

Se puede observar que en ambos problemas, las funciones objetivo son funciones de valor intervalo. Además, el programa matemático (IP2) se puede expresar como el problema (IP1) considerando el conjunto admisible
$X = \{x \in \mathbb{R}^n / g_i(x) \leq 0, i = 1, \ldots, m\} \subset \mathbb{R}^n$,
donde g_i, $i = 1, \ldots, m$ son funciones de valor real definidas en $\mathbb{R}^n$. Para poder resolver estos programas matemáticos previamente se tiene que proponer una relación de orden en $\mathscr{I}$. A continuacion se proponen tres tipos de relaciones de orden en $\mathscr{I}$

Definición 4.1. *(Wu [9]) Sean* $C = [c^I, c^S]$, $D = [d^I, d^S]$, *elementos de* $\mathscr{I}$. *Se tiene que*

$$C \preceq_{\text{IS}} D \text{ siempre y cuando, } c^I \leq d^I \text{ y } c^S \leq d^S \quad (5)$$

Ejemplo 7. *sean* $C = [3, 5]$ *y* $D = [4, 5]$, *se observa que* $C \preceq_{\text{IS}} D$

Se puede observar que "$\preceq_{\text{IS}}$"es relación de orden parcial sobre $\mathscr{I}$, porque es reflexiva, transitiva y antisimétrica.
Además, se escribe $C \prec_{\text{IS}} D$, siempre y cuando, $C \preceq_{\text{IS}} D$ y $C \neq D$. Equivalentemente, $C \prec_{\text{IS}} D$, cuando y solo cuando,

$$\begin{cases} c^I < d^I \\ c^S \leq d^S \end{cases} \text{ o } \begin{cases} c^I \leq d^I \\ c^S < d^S \end{cases} \text{ o } \begin{cases} c^I < d^I \\ c^S < d^S \end{cases}. \quad (6)$$

Definición 4.2. *(Wu [9]) Sea* $\mathbf{x}^*$ *una solución factible, esto es,* $\mathbf{x}^* \in X$. *Se dice que* $\mathbf{x}^*$ *es una* solución de tipo I *del problema (IP1) si no existe* $\bar{\mathbf{x}} \in X$ *de modo que* $f(\bar{\mathbf{x}}) \prec_{\text{IS}} f(\mathbf{x}^*)$.

A continuación se presentará la segunda relación de orden parcial (presentada por Ishibuchi y Tanaka [5]). Sea $A = [a^I, a^S] \in \mathscr{I}$. Se puede calcular su centro $a^C = \frac{1}{2}(a^I + a^S)$ y su semilongitud $a^R = \frac{w(A)}{2} = \frac{1}{2}(a^S - a^I)$ de A. En esta situación se puede emplear la notación $\langle a^C, a^R \rangle$ para denotar el intervalo cerrado A como $A = \langle a^C, a^R \rangle$, esto es, $A = [a^I, a^S] = \langle a^C, a^R \rangle$.

Definición 4.3. *Sean* $A = [a^I, a^S] = \langle a^C, a^R \rangle$, $B = [b^I, b^S] = \langle b^C, b^R \rangle \in \mathscr{I}$. *Se dice que*

$$A \preceq_{\text{CR}} B \text{ siempre y cuando, } a^C \leq b^C \text{ y } a^R \leq b^R \quad (7)$$

También, se escribe $A \prec_{\text{CR}} B$ cuando y solo cuando $A \preceq_{\text{CR}} B$ y $A \neq B$. Además, se escribe equivalentemente, $A \prec_{\text{CR}} B$ siempre y cuando,

$$\begin{cases} a^C < b^C \\ a^R \leq b^R \end{cases} \text{ o } \begin{cases} a^C \leq b^C \\ a^R < b^R \end{cases} \text{ o } \begin{cases} a^C < b^C \\ a^R < b^R \end{cases}. \quad (8)$$

Definición 4.4. *Sea* $\mathbf{x}^*$ *una solución factible, es decir,* $\mathbf{x}^* \in X$. *Decimos que* $\mathbf{x}^*$ *es una* solución de tipo II *del problema (IP1) si no existe* $\bar{\mathbf{x}} \in X$ *de modo que* $f(\bar{\mathbf{x}}) \prec_{\text{IS}} f(\mathbf{x}^*)$ *o* $f(\bar{\mathbf{x}}) \prec_{\text{CR}} f(\mathbf{x}^*)$.

Observación 4.1. Sea $\mathbf{x}^*$ una solución factible, es decir, $\mathbf{x}^* \in X$. Se puede observar, por la propia definición 4.4, de que si $\mathbf{x}^*$ es una solución de tipo I del problema (IP1), entonces $\mathbf{x}^*$ es también una solución de tipo II del problema (IP1).

Finalmente se tiene la tercera relación de orden presentada por Chalco [3]. Sea $A = [a^I, a^S] \in \mathscr{I}$ su longitud o ancho es $w(A)$ y se denotará por a^W. Así se tiene que

$$a^W = w(A) = a^S - a^I$$

Definición 4.5. *Sean* $A = [a^I, a^S]$, $B = [b^I, b^S] \in \mathscr{I}$. *Se tiene que*

$$A \preceq_{\text{IW}} B \text{ siempre y cuando, } a^I \leq b^I \text{ y } a^W \leq b^W \quad (9)$$

Se puede observar que "$\preceq_{\text{IW}}$"es una relación de orden parcial sobre $\mathscr{I}$, porque es reflexiva, transitiva y antisimétrica.
Además, se escribe $A \prec_{\text{IW}} B$ siempre y cuando, $A \preceq_{\text{IW}} B$ y $A \neq B$. Equivalentemente, $A \prec_{\text{IW}} B$ si, y solo si,

$$\begin{cases} a^I < b^I \\ a^W \leq b^W \end{cases} \text{ o } \begin{cases} a^I \leq b^I \\ a^W < b^W \end{cases} \text{ o } \begin{cases} a^I < b^I \\ a^W < b^W \end{cases}. \quad (10)$$

Definición 4.6. *(Chalco [3]) Sea* $\mathbf{x}^*$ *una solución factible, esto es,* $\mathbf{x}^* \in X$. *Decimos que* $\mathbf{x}^*$ *es una* solución de tipo III *del problema (IP1) si no existe* $\bar{\mathbf{x}} \in X$ *tal que* $f(\bar{\mathbf{x}}) \prec_{\text{IW}} f(\mathbf{x}^*)$.

Proposición 4.1. *Sean* $A, B \in \mathscr{I}$ *cualesquiera. Si* $A \preceq_{IW} B$, *entonces* $A \preceq_{IS} B$.

Demostración. Ya que A y B son intervalos cerrados, de modo que $A \preceq_{IW} B$, se tiene

$$a^I \leq b^I \quad \text{y} \quad a^S - a^I = w(A) \leq w(B) = b^S - b^I.$$

Así,

$$a^S - a^I + b^I \leq b^S$$

entonces

$$a^S \leq a^S + (b^I - a^I) \leq b^S.$$

En consecuencia, $A \preceq_{IS} B$. □

Se debe notar que el recíproco de la Proposición (4.1) no es válida. Por ejemplo, si consideramos $A = [-2, 0]$ y $B = [-1, 0]$, entonces $A \preceq_{IS} B$ pero $A \not\preceq_{IW} B$.

Teorema 4.2. *Sea* $\mathbf{x}^*$ *una solución factible de (IP1). Si* $\mathbf{x}^*$ *es una solución tipo I del problema (IP1), entonces* $\mathbf{x}^*$ *es una solución del tipo III del problema (IP1).*

Demostración. como $\mathbf{x}^* \in X$. Suponer que $\mathbf{x}^*$ no es una solución tipo III del problema (IP1), entonces existe $\mathbf{x} \in X$ tal que $F(\mathbf{x}) \preceq_{IW} F(\mathbf{x}^*)$ y $F(\mathbf{x}) \neq F(\mathbf{x}^*)$. de la Proposición (4.1) $F(\mathbf{x}) \preceq_{IS} F(\mathbf{x}^*)$ y $F(\mathbf{x}) \neq F(\mathbf{x}^*)$, lo cual es una contradicción, generado por lo supuesto. □

A continuación, se redefinirá el concepto de funciones convexas desde la perspectiva de funciones de valor intervalo.

Definición 4.7. *(Chalco [3]) Sea el conjunto convexo* $X \subseteq \mathbb{R}^n$ *y la función* $F : X \to \mathscr{I}$ *de valor intervalo, con* $F(\mathbf{x}) = [F^I(\mathbf{x}), F^S(\mathbf{x})]$. *Se dice que* F *es*

(i) IS-convexa *en* $\mathbf{x}^*$ *si*

$$F(\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \preceq_{\text{IS}} \lambda F(\mathbf{x}^*) + (1-\lambda) f(\mathbf{x}), \quad (11)$$

para cada $\lambda \in]0, 1[$ *y cada* $\mathbf{x} \in X$

(ii) CR-convexa *en* $\mathbf{x}^*$ *si*

$$F(\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \preceq_{\text{CR}} \lambda F(\mathbf{x}^*) + (1-\lambda) f(\mathbf{x}), \quad (12)$$

para cada $\lambda \in]0, 1[$ *y cada* $\mathbf{x} \in X$

(iii) IW-convexa *en* $\mathbf{x}^*$ *si*

$$F(\lambda\mathbf{x}^*+(1-\lambda)\mathbf{x}) \preceq_{\text{IW}} \lambda F(\mathbf{x}^*)+(1-\lambda)f(\mathbf{x}), \quad (13)$$

para cada $\lambda \in]0,1[$ *y cada* $\mathbf{x} \in X$

Proposición 4.3. *Sean* X *un subconjunto convexo de* $\mathbb{R}^n$ *y* $F : X \to \mathscr{I}$ *una función de valor intervalo, con* $F(\mathbf{x}) = [F^I(\mathbf{x}), F^S(\mathbf{x})]$. *Se cumple que:*

(i) F *es IS-convexa en* $\mathbf{x}^*$ *siempre y cuando* F^I *y* F^S *son convexas en* $\mathbf{x}^*$.

(ii) F *es CR-convexa en* $\mathbf{x}^*$ *siempre y cuando* F^C *y* F^R *son convexas en* $\mathbf{x}^*$.

(iii) F *es IW-convexa en* $\mathbf{x}^*$ *siempre y cuando* F^I *y* F^W *son convexas en* $\mathbf{x}^*$.

(iv) Si F *es IW-convexa en* $\mathbf{x}^*$, *entonces* F *es también IS-convexa en* $\mathbf{x}^*$.

Demostración. Los incisos (i), (ii) y (iii) son consecuencia de la definición, y (iv) se obtiene como consecuencia de la proposición 4.1. □

Ahora, se extenderá estas relaciones de orden a los vectores cuyos componentes son valor intervalo.

Definición 4.8. *Sea* $C = (C_1, C_2, \ldots, C_q)$ *es denominado vector de valor intervalo si* $C_j \in \mathscr{I}$, *para cualquier* $j = 1, 2, \ldots, q$

Definición 4.9. *Sean* $C = (C_1, C_2, \ldots, C_q)$ *y* $D = (D_1, D_2, \ldots, D_q)$ *vectores de valor intervalo, se dice que*

(i) $C \preceq_{\text{IS}} D$ *siempre y cuando* $C_j \preceq_{\text{IS}} D_j$ *con* $j = 1, 2, \ldots, q$.

(ii) $C \prec_{\text{IS}} D$ *siempre y cuando,* $C_j \preceq_{\text{IS}} D_j$, *para* $j = 1, 2, \ldots, q$; *y* $C_k \prec_{\text{IS}} D_k$ *para al menos un índice* k.

Definición 4.10. *Sean* $C = (C_1, C_2, \ldots, C_q)$ *y* $D = (D_1, D_2, \ldots, D_q)$ *vectores de valor intervalo se dice que*

(i) $C \preceq_{\text{CR}} D$ *siempre y cuando* $C_j \preceq_{\text{CR}} D_j$, *con* $j = 1, 2, \ldots, q$.

(ii) $C \prec_{\text{CR}} D$ *siempre y cuando* $C_j \preceq_{\text{CR}} D_j$, *para* $j = 1, 2, \ldots, q$; *y* $C_k \prec_{\text{CR}} D_k$ *para al menos un índice* k.

Definición 4.11. *Sean* $C = (C_1, C_2, \ldots, C_q)$ *y* $D = (D_1, D_2, \ldots, D_q)$ *vectores de valor intervalo se dice que*

(i) $C \preceq_{\text{IW}} D$ *siempre y cuando* $C_j \preceq_{\text{IW}} D_j$, *con* $j = 1, 2, \ldots, q$.

(ii) $C \prec_{\text{IW}} D$ *siempre y cuando* $C_j \preceq_{\text{IW}} D_j$, *para* $j = 1, 2, \ldots, q$; *y* $C_k \prec_{\text{IW}} D_k$ *para al menos un índice* k.

Proposición 4.4. *Sean* $C = (C_1, C_2, \ldots, C_q)$ *y* $D = (D_1, D_2, \ldots, D_q)$ *vectores de valor intervalo se dice que*

(i) si $C \preceq_{\text{IW}} D$, *entonces* $C \preceq_{\text{IS}} D$.

(ii) si $C \prec_{\text{IW}} D$, *entonces* $C \prec_{\text{IS}} D$.

Demostración. (i) Como C y D son vectores valor intervalo y $C \preceq_{\text{IW}} D$, entonces $C_j \preceq_{\text{IW}} D_j$, para $j = 1, 2, \ldots, q$.
Luego, por la proposición (4.1) se tiene que $C_j \preceq_{\text{IS}} D_j$, $j = 1, 2, \ldots, q$.
En consecuencia, se cumple que $C \preceq_{\text{IS}} D$

(ii) análogo a lo anterior.

□

Definición 4.12. *Sea el conjunto convexo* $X \subseteq \mathbb{R}^n$ *y la función* F *multivalor intervalo definida sobre* X, *esto es,* $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$, *donde cada una de las* $F_i : X \to \mathscr{I}$ *son funciones de valor intervalo, para* $i = 1, 2, \ldots, q$. *Se dice que* F *es*

(I) IS-convexa *en* $\mathbf{x}^*$ *si cada* F_j *es* IS-convexa *en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$.

(II) CR-convexa *en* $\mathbf{x}^*$ *si cada* F_j *es* CR-convexa *en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$.

(III) IW-convexa *en* $\mathbf{x}^*$ *si cada* F_j *es* IW-convexa *en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$.

Proposición 4.5. *Sea el conjunto convexo* $X \subseteq \mathbb{R}^n$ *y la función* F *multivalor intervalo definida sobre* X, *esto es,* $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$, *donde cada una de las* $F_j : X \to \mathscr{I}$ *son funciones de valor intervalo, para* $j = 1, 2, \ldots, q$. *Se cumple que:*

(I) F *es IS-convexa en* $\mathbf{x}^*$ *siempre y cuando* F_j^I *y* F_j^S *son convexas en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$..

(II) F *es CR-convexa en* $\mathbf{x}^*$ *siempre y cuando* F_j^C *y* F_j^R *son convexas en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$..

(III) F *es IW-convexa en* $\mathbf{x}^*$ *siempre y cuando* F_j^I *y* F_j^W *son convexas en* $\mathbf{x}^*$, $j = 1, 2, \ldots, q$.

(iv) Si F *es IW-convexa en* $\mathbf{x}^*$, *entonces* F *es también IS-convexa en* $\mathbf{x}^*$.

5. Optimizacion de programas matemáticos con función multiobjetivo de valor intervalo

Las funciones multivalor intervalo F, las cuales están definidas sobre un abierto $X \subset \mathbb{R}^n$, son aquellas de la forma $F(t) = (F_1(t), F_2(t), \ldots, F_q(t))$, $t \in X$, donde cada una de las $F_i : X \to \mathscr{I}$ son funciones de valor intervalo, para $i = 1, 2, \ldots, q$, esto es, $F_i(t) = [F_i^I(t), F_i^S(t)]$, $t \in X$, para $i = 1, 2, \ldots, q$. A continuación se presentan los programas matemáticos, cuyas funciones multiobjetivo son funciones multivalor intervalo

$$\textbf{(MP1)} \quad \begin{array}{ll} \text{mín} & F(x) = (F_1(x), \ldots, F_q(x)) \\ \text{sujeto a} & x = (x_1, \ldots, x_n) \in X \subseteq \mathbb{R}^n. \end{array}$$

donde los $F_i(x) = [F_i^I(x), F_i^S(x)]$, $x \in X$, para $i = 1, 2, \ldots, q$, son funciones de valor intervalo y el conjunto $X \subset \mathbb{R}^n$ será considerado convexo en $\mathbb{R}^n$.

$$\textbf{(MP2)} \quad \begin{array}{ll} \text{mín} & F(x) = (F_1(x), \ldots, F_q(x)) \\ \text{sujeto a} & g_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

donde las funciones restricción de valor real $g_i : \mathbb{R}^n \to \mathbb{R}$ son convexas en $\mathbb{R}^n$ para $i = 1, \ldots, m$; y las componentes F son funciones de valor intervalo. Además, el programa matemático (MP2) se puede expresar como el problema (MP1) considerando el conjunto admisible $X = \{x \in \mathbb{R}^n / g_i(x) \leq 0, i = 1, \ldots, m\} \subset \mathbb{R}^n$.

5.1. Condiciones de optimalidad de Pareto

Se presentan los diferentes conceptos de solución óptima de Pareto.

Definición 5.1. *Sea $\boldsymbol{x}^*$ una solución factible del problema (MP1), se dice que $\boldsymbol{x}^*$*

(i) es una solución óptima de Pareto tipo-I del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \prec_{IS} F(\boldsymbol{x}^*)$.

(ii) es una solución óptima de Pareto fuertemente tipo-I del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \preceq_{IS} F(\boldsymbol{x}^*)$.

(iii) es una solución óptima de Pareto débilmente tipo-I del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F_k(\overline{\boldsymbol{x}}) \prec_{IS} F_k(\boldsymbol{x}^*)$ para todo $k = 1, \ldots, q$.

Observación 5.1. *Sea $X_{wp}^{(I)}$, $X_p^{(I)}$ y $X_{sp}^{(I)}$, denotan el conjunto de todas las soluciones débilmente óptimas de Pareto tipo-I, soluciones óptimas de Pareto tipo-I y soluciones fuertemente óptimas de Pareto tipo-I, respectivamente. Se puede observar que que $X_{sp}^{(I)} \subseteq X_p^{(I)} \subseteq X_{wp}^{(I)}$.*

Definición 5.2. *Sea $\boldsymbol{x}^*$ una solución factible del problema (MP1), se dice que $\boldsymbol{x}^*$*

(i) es una solución óptima de Pareto tipo-II del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \prec_{CR} F(\boldsymbol{x}^*)$.

(ii) es una solución óptima de Pareto fuertemente tipo-II del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \preceq_{CR} F(\boldsymbol{x}^*)$.

(iii) es una solución óptima de Pareto débilmente tipo-II del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F_k(\overline{\boldsymbol{x}}) \prec_{CR} F_k(\boldsymbol{x}^*)$ para todo $k = 1, \ldots, q$.

Observación 5.2. *Sea $X_{wp}^{(II)}$, $X_p^{(II)}$ y $X_{sp}^{(II)}$, denotan el conjunto de todas las soluciones débilmente óptimas de Pareto tipo-II, soluciones óptimas de Pareto tipo-II y soluciones fuertemente óptimas de Pareto tipo-II, respectivamente. Se puede observar que que $X_{sp}^{(II)} \subseteq X_p^{(II)} \subseteq X_{wp}^{(II)}$.*

Definición 5.3. *Sea $\boldsymbol{x}^*$ una solución factible del problema (MP1), se dice que $\boldsymbol{x}^*$*

(i) es una solución óptima de Pareto tipo-III del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \prec_{IW} F(\boldsymbol{x}^*)$.

(ii) es una solución óptima de Pareto fuertemente tipo-III del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F(\overline{\boldsymbol{x}}) \preceq_{IW} F(\boldsymbol{x}^*)$.

(iii) es una solución óptima de Pareto débilmente tipo-III del problema (MP1) si no existe $\overline{\boldsymbol{x}} \in X$ de modo que $F_k(\overline{\boldsymbol{x}}) \prec_{IW} F_k(\boldsymbol{x}^*)$ para todo $k = 1, \ldots, q$.

Observación 5.3. *Sea $X_{wp}^{(III)}$, $X_p^{(III)}$ y $X_{sp}^{(III)}$, denotan el conjunto de todas las soluciones débilmente óptimas de Pareto tipo-III, soluciones óptimas de Pareto tipo-III y soluciones fuertemente óptimas de Pareto tipo-III, respectivamente. Se puede observar que que $X_{sp}^{(III)} \subseteq X_p^{(III)} \subseteq X_{wp}^{(III)}$.*

Teorema 5.1. *Sea X un conjunto admisible de (MP1). Se cumple*

(i) $X_{SP}^{(I)} \subset X_{SP}^{(III)}$

(ii) $X_P^{(I)} \subset X_P^{(III)}$

(iii) $X_{WP}^{(I)} \subset X_{WP}^{(III)}$

Demostración. (i) considerando que $\mathbf{x}^* \in X_{SP}^{(I)}$
Por contradicción, suponer que $\mathbf{x}^* \notin X_{SP}^{(III)}$.
Luego, por definición 5.3 existe $\overline{\mathbf{x}} \in X$ de modo que $F(\overline{\mathbf{x}}) \preceq_{IW} F(\mathbf{x}^*)$. Por la proposición (4.4), se tiene que $F(\overline{\mathbf{x}}) \preceq_{IS} F(\mathbf{x}^*)$, lo cual es una contradicción. En consecuencia, $X_{SP}^{(I)} \subset X_{SP}^{(III)}$

(ii) similar a (i)

(iii) considerando que $\mathbf{x}^* \in X_{WP}^{(I)}$. Por contradicción, suponer que $\mathbf{x}^* \notin X_{WP}^{(III)}$.
Luego, por definición 5.3, existe $\overline{\mathbf{x}} \in X$ de modo que $F_k(\overline{\mathbf{x}}) \prec_{IW} F_k(\mathbf{x}^*)$, para $k = 1, 2, \ldots, q$
de la proposición (4.4), $F_k(\overline{\mathbf{x}}) \prec_{IS} F_k(\mathbf{x}^*)$, para $k = 1, 2, \ldots, q$; lo cual es una contradicción. En consecuencia, $X_{WP}^{(I)} \subset X_{WP}^{(III)}$.

□

5.2. Condiciones de optimalidad de Karush Khun Tucker

Sea el siguiente programa matemático

$$(P) \quad \begin{array}{ll} min & f(\mathbf{x}) = f(x_1, \ldots, x_n) \\ \text{sujeto a} & g_i(\mathbf{x}) \leqslant 0, \quad i = 1, \ldots, m \end{array}$$

donde f y g_i son funciones reales definidas en $\mathbb{R}^n$, $i = 1, \ldots, n$. Suponga que las funciones restricción g_i son convexas en $\mathbb{R}^n$ para cada $i = 1, \ldots, m$, en consecuencia el conjunto factible $X = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leqslant 0, i = 1, \ldots, m\}$ es un subconjunto convexo de $\mathbb{R}^n$. La conocida condición Karush-Kuhn-Tucker (ver Bazaraa [2]), para el problema (P), es declarada de la siguiente manera:

Teorema 5.2. *[2] Dado el problema (P). Asumiendo que las restricciones $g_i : \mathbb{R}^n \to \mathbb{R}$ son funciones convexas en $\mathbb{R}^n$ para $i = 1, \ldots, m$, $X = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \leqslant 0, i = 1, \ldots, m\}$ es un conjunto admisible, $\boldsymbol{x}^* \in X$, la función objetivo $f : \mathbb{R}^n \to \mathbb{R}$ es convexa, f y g_i son continuamente diferenciable en $\boldsymbol{x}^*$, $i = 1, \ldots, m$. Si existen los multiplicadores de Lagrange $0 \leqslant \mu_i \in \mathbb{R}$, $i = 1, \ldots, m$, de modo que*

(I) $\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) = 0$;

(II) $\mu_i g_i(\boldsymbol{x}^*) = 0$ *para todo* $i = 1, \ldots, m$.

entonces $\boldsymbol{x}^*$ es una solución óptima de (P)

Definición 5.4. *[9] Se dice que las funciones de restricción del problema (MP2) satisfacen las condiciones KKT en $\boldsymbol{x}^*$ si estas son convexas en $\mathbb{R}^n$ y continuamente diferenciables en $\boldsymbol{x}^*$.*

Ahora, las condiciones Karush Khun Tucker (KKT) presentadas en el teorema 5.2 se ampliarán para el caso de las funciones de valor intervalo (Wu[9], Chalco [3]) así como para las funciones multivalor intervalo.

Teorema 5.3. *Dado el programa matemático (MP2), $x^* \in X$. Considerando que F es continuamente gH-diferenciable en x^*; F_j^I, F_j^S son funciones convexas, para $j = 1, \ldots, q$. Si existen los multiplicadores de Lagrange $0 < \lambda_j \in \mathbb{R}$, $j = 1, \ldots, q$ y $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, de modo que las siguientes condiciones KKT se cumplan:*

(I) $\sum_{j=1}^{q} \lambda_j \nabla(F_j^I + F_j^S)(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$ *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_P^{(I)}$ y $\boldsymbol{x}^* \in X_P^{(III)}$ para el programa matemático (MP2).

Demostración. Como F es continuamente gH-diferenciable en $\mathbf{x}^*$, debido a la proposición 2.9 se tiene que $(F_j^I + F_j^S)$ es continuamente diferenciable en $\mathbf{x}^*$ para $j = 1, \ldots, q$.
Definiendo la función f por $f(x) = \sum_{j=1}^{q} \lambda_j \nabla(F_j^I + F_j^S)(x)$
Como F_j^I y F_j^S, para $j = 1, \ldots, q$, son funciones reales convexas, se tiene que f es convexa y continuamente diferenciable en $\mathbf{x}^*$.
Así $\nabla f(\mathbf{x}^*) = \sum_{j=1}^{q} \lambda_j \nabla(F_j^I + F_j^S)(\mathbf{x}^*)$ y con las condiciones (i) y (ii), se tendría que:

(I)' $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*) = 0$

(II)' $\mu_j g_j(\mathbf{x}^*) = 0$ para todo $j = 1, \ldots, m$.

Aplicando el teorema 5.2 (condiciones KKT a funciones reales), se tiene que $\mathbf{x}^*$ es una solución óptima de la función real f.
Suponiendo que $\mathbf{x}^* \notin X_P^{(I)}$, esto significa que hay un $x \in X$ y $1 \leq k \leq q$ de modo que $F_k(\overline{\mathbf{x}}) \prec_{IS} F_k(\mathbf{x}^*)$, esto implica que $f(x) < f(\mathbf{x}^*)$, lo cual contradice de que f presente un valor óptimo en $\mathbf{x}^*$. En consecuencia, se tiene que $\mathbf{x}^* \in X_P^{(I)}$ y por la proposición 5.1, se tiene que $\mathbf{x}^* \in X_P^{(III)}$ □

Corolario 5.4. *Dado el programa matemático (MP2), $x^* \in X$. Considerando que F es continuamente gH-diferenciable en x^* e IS-convexa en $\boldsymbol{x}^*$. Si existen los multiplicadores de Lagrange $0 < \lambda_j \in \mathbb{R}$, $j = 1, \ldots, q$ y $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, de modo que las siguientes condiciones KKT se cumplan:*

(I) $\sum_{j=1}^{q} \lambda_j \nabla(F_j^I + F_j^S)(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$ *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_P^{(I)}$ y $\boldsymbol{x}^* \in X_P^{(III)}$ para el programa matemático (MP2).

Demostración. Como F es IS-convexa en $\mathbf{x}^*$, aplicando la proposición 4.3, se tiene que F_j^I, F_j^S son funciones convexas, para $j = 1, \ldots, q$. Asimismo como F es gH-diferenciable en $\mathbf{x}^*$, aplicando la proposición 2.9 entonces, se tiene que $(F_j^I + F_j^S)$ son continuamente diferenciables en $\mathbf{x}^*$, para $j = 1, \ldots, q$.
Definiendo la función f por $f(x) = \sum_{j=1}^{q} \lambda_j \nabla(F_j^I + F_j^S)(x)$
se tiene que f es convexa y continuamente diferenciable en $\mathbf{x}^*$ y por el teorema anterior se obtendría el resultado esperado. □

Ejemplo 8. *sea la función F multivalor intervalo dada por $F(x) = (F_1(x), F_2(x)$ donde $F_1(x) = [-|(x-1)|, |x-1|]$*
$F_2(x) = [-|\frac{(x-1)}{2}|, |\frac{(x-1)}{2}|]$
Se tiene el siguiente programa matemático

$$\begin{array}{lll} \text{mín} & F(x) = (F_1(x), F_2(x)) & \\ & x - 2 & \leq 0 \\ \textit{sujeto a} & -x & \leq 0 \end{array}$$

F es continuamente gH-diferenciable y las condiciones del teorema 5.3 son satisfechas en $x = 1$

Teorema 5.5. *Dado el programa matemático (MP2). Considerando que F es CR-convexa y (debil) continuamente diferenciable en $\boldsymbol{x}^*$. Si existen los multiplicadores de Lagrange*
$0 < \lambda_j^C, \lambda_j^R \in \mathbb{R}$, $j = 1, \ldots, q$ y $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, de modo que las siguientes condiciones KKT se cumplan:

(I) $\sum_{j=1}^{q} \lambda_j^C \nabla F_j^C(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j^R \nabla F_j^R(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$ *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_P^{(II)}$ para el programa matemático (MP2).

Demostración. Como $F = (F_1, F_2, \ldots, F_q))$ es una función multivalor intervalo, esto es $F_j(x) = (F_j^I, F_j^S)$, (debil) continuamente diferenciable y convexa, entonces los F_j^C y F_j^R son continuamente diferenciables (por definición 2.8 y proposición 2.9) y convexas (proposición 4.3) en $\mathbf{x}^*$ para $j = 1, 2, \ldots, q$.
Definiendo la función f por

$$f(x) = \sum_{j=1}^{q} \lambda_j^C F_j^C(\mathbf{x}^*) + \sum_{j=1}^{q} \lambda_j^R F_j^R(\mathbf{x}^*)$$

Como F_j^C y F_j^R, para $j = 1, \ldots, q$, son funciones reales continuamente diferenciables y convexas, se tiene que f es convexa y continuamente diferenciable en $\mathbf{x}^*$.
Así $\nabla f(\mathbf{x}^*) = \sum_{j=1}^{q} \lambda_j^C \nabla(F_j^C)(\mathbf{x}^*) + \sum_{j=1}^{q} \lambda_j^R \nabla(F_j^R)(\mathbf{x}^*)$ y con las condiciones (i) y (ii), se tendría que:

(I') $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*) = 0$

(II') $\mu_j g_j(\mathbf{x}^*) = 0$ para todo $j = 1, \ldots, m$.

Aplicando el teorema 5.2 (condiciones KKT a funciones reales), se tiene que $\mathbf{x}^*$ es una solución óptima de la función real objetivo f bajo las restricciones de (MP2).
Suponiendo que $\mathbf{x}^* \notin X_P^{(II)}$, esto significa que hay un $\overline{\mathbf{x}} \in X$ y $1 \leq k \leq q$ de modo que $F_k(\overline{\mathbf{x}}) \prec_{CR} F_k(\mathbf{x}^*)$, esto implica que $f(\overline{\mathbf{x}}) < f(\mathbf{x}^*)$, lo cual contradice de que f presente un valor óptimo en $\mathbf{x}^*$. En consecuencia, se tiene que $\mathbf{x}^* \in X_P^{(II)}$ □

Teorema 5.6. *Dado el programa matemático (MP2). Considerando que F es IW-convexa y (debil) continuamente diferenciable en $\boldsymbol{x}^*$. Si existen los multiplicadores de Lagrange*
$0 < \lambda_j^I, \lambda_j^W \in \mathbb{R}$, $j = 1, \ldots, q$ *y* $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, *de modo que las siguientes condiciones KKT se cumplan:*

(I) $\sum_{j=1}^{q} \lambda_j^I \nabla F_j^I(\boldsymbol{x}^*) + \sum_{j=1}^{q} \lambda_j^W \nabla F_j^W(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$ *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_P^{(III)}$ *para el programa matemático (MP2).*

Demostración. Como $F = (F_1, F_2, \ldots, F_q)$ es una función multivalor intervalo, esto es $F_j(x) = (F_j^I, F_j^W)$, (debil) continuamente diferenciable e IW- convexas en $\mathbf{x}^*$ para $j = 1, 2, \ldots, q.$, entonces los F_j^I, F_j^W son continuamente diferenciables (por definición 2.8 y proposición 2.9) y convexas (proposición 4.3) en $\mathbf{x}^*$ para $j = 1, 2, \ldots, q$.
Definiendo la función f por $f(x) = \sum_{j=1}^{q} \lambda_j^I F_j^I(x) + \sum_{j=1}^{q} \lambda_j^W F_j^W(x)$
Como F_j^I, F_j^W (para $j = 1, \ldots, q$), y g_i (para $j = 1, \ldots, m$) , son funciones reales continuamente diferenciables y convexas, se tiene que f es convexa y continuamente diferenciable en $\mathbf{x}^*$.
Así $\nabla f(\mathbf{x}^*) = \sum_{j=1}^{q} \lambda_j^I \nabla(F_j^I)(\mathbf{x}^*) + \sum_{j=1}^{q} \lambda_j^W \nabla(F_j^W)(\mathbf{x}^*)$ y con las condiciones (i) y (ii), se tendría que:

(I') $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*) = 0$

(II') $\mu_j g_j(\mathbf{x}^*) = 0$ para todo $j = 1, \ldots, m$.

Aplicando el teorema 5.2 (condiciones KKT a funciones reales), se tiene que $\mathbf{x}^*$ es una solución óptima de la función real objetivo f bajo las restricciones de (MP2).
Suponiendo que $\mathbf{x}^* \notin X_P^{(III)}$, esto significa que hay un $\overline{\mathbf{x}} \in X$ y $1 \leq k \leq q$ de modo que $F_k(\overline{\mathbf{x}}) \prec_{IW} F_k(\mathbf{x}^*)$, esto implica que $f(x) < f(\mathbf{x}^*)$, lo cual contradice de que f presente un valor óptimo en $\mathbf{x}^*$. En consecuencia, se tiene que $\mathbf{x}^* \in X_P^{(III)}$ del programa matemático MP2) □

Aplicación numérica
Sea la función F multivalor intervalo dada por

$$F(x_1, x_2) = (F_1(x_1, x_2), F_2(x_1, x_2))$$

donde sus componentes son funciones de valor intervalo dadas por
$F_1(x_1, x_2) = [x_1^2 + 2x_1 + x_2^2 - 2x_2 + 3, x_1^2 + 2x_1 + x_2^2 - 2x_2 + 4]$
$F_2(x_1, x_2) = [2x_1^2 + 4x_1 + 2x_2^2 - 4x_2 + 7, 2x_1^2 + 4x_1 + 2x_2^2 - 4x_2 + 7]$
Se tiene el siguiente programa matemático (tipo MP2)

$$\begin{aligned} min \quad & (F(x_1, x_2) = F_1(x_1, x_2), F_2(x_1, x_2)) \\ \text{sujeto a} \quad & -x_1 - x_2 + 1 \leqslant 0 \\ & -3x_1 - x_2 + 4 \leqslant 0, \\ & -x_1 - 1 \leqslant 0, \\ & -x_2 + 1 \leqslant 0. \end{aligned}$$

Entonces tenemos las funciones de valor intervalo, componentes de F

$$F_1^I(\mathbf{x}) = x_1^2 + 2x_1 + x_2^2 - 2x_2 + 3$$
$$F_1^W(\mathbf{x}) = 1$$
$$F_2^I(\mathbf{x}) = 2x_1^2 + 4x_1 + 2x_2^2 - 4x_2 + 7$$
$$F_2^W(\mathbf{x}) = 1$$

y las funciones restricción:

$$g_1(\mathbf{x}) = -x_1 - x_2 + 1, \quad g_2(\mathbf{x}) = -3x_1 - x_2 + 4,$$
$$g_3(\mathbf{x}) = -x_1 - 1, \quad g_4(\mathbf{x}) = -x_2.$$

Se puede ver que las funciones de valor intervalo asi como las funciones restricción satisfacen las condiciones del Teorema .
Para ver que cumplan las condiciones KKT del teorema se tiene la siguiente expresión:

$$\lambda_1^I \begin{bmatrix} 2x_1 + 2 \\ 2x_2 - 2 \end{bmatrix} + \lambda_2^I \begin{bmatrix} 4x_1 + 4 \\ 4x_2 - 4 \end{bmatrix} + \lambda_1^W \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \lambda_2^W \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \mu_1 \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \mu_2 \begin{bmatrix} -3 \\ -1 \end{bmatrix} + \mu_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \mu_4 \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

esto es, se tiene que resolver las siguientes ecuaciones simultáneas:

$$\begin{cases} \lambda_1^I(2x_1+2)+\lambda_2^I(4x_1+4)-\mu_1-3\mu_2-\mu_3=0, \\ \lambda_1^I 2x_2-2()+\lambda_2^I(4x_2-4)-\mu_1-\mu_2-\mu_4=0. \end{cases}$$

Resolviendo, se obtiene

$$(x_1^*, x_2^*) = (4/5, 8/5)$$

$$\lambda_1^I = 1/2; \quad \lambda_2^I = 1/4$$

$$\mu_1 = \mu_3 = \mu_4 = 0 \text{ y } \mu_2 = 6/5$$

Como $\mu_i g_i(\mathbf{x}^*) = \mu_i g_i(9/5, 3/5) = 0$ para $i = 1, \ldots, 4$. Se concluye que $(x_1^*, x_2^*) = (4/5, 8/5)$ es una solución optima Pareto tipo-II.

Teorema 5.7. *Dado el programa matemático (MP2), $x^* \in X$. Considerando que la función multiobjetivo F tiene alguna componente F_j (función valor intervalo) IS-convexa y continuamente gH diferenciable en x^*, para algún $j \in \{1, 2, \ldots, q\}$. Si existen los multiplicadores de Lagrange $0 < \lambda \in \mathbb{R}$ y $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, de modo que las siguientes condiciones KKT se cumplan:*

(I) $\lambda \nabla(F_j^I + F_j^S)(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$ *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_{WP}^{(I)}$ *para el programa matemático (MP2).*

Demostración. Como $F = (F_1, F_2, \ldots, F_q))$ es una función multivalor intervalo, esto es $F_j(x) = [F_j^I(x), F_j^S(x)]$ son funciones de valor intervalo, para $j = 1, 2. \ldots, q$.
Considerando que alguna de la funciones componentes F_j, para algún
$j \in \{1, 2. \ldots, q\}$, es IS-convexa y continuamente gH diferenciable en x^*, condiciones del teorema.
Definiendo la función f por $f(x) = \lambda(F_j^I + F_j^S)(x)$
Como F_j es IS-convexa $\mathbf{x}^*$, por la proposición (4.1) se tiene que F_j^I y F_j^S, son funciones reales convexas, se tiene que f es convexa. Asimismo F_j es continuamente gH-diferenciable $\mathbf{x}^*$, por la proposición 2.9, entonces $(F_j^I + F_j^S)$ es continuamente diferenciable en $\mathbf{x}^*$.
Así $\nabla f(\mathbf{x}^*) = \nabla(F_j^I + F_j^S)(\mathbf{x}^*)$ y con las condiciones (i) y (ii), se tendría que:

(I)' $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*) = 0$

(II)' $\mu_j g_j(\mathbf{x}^*) = 0$ para todo $j = 1, \ldots, m$.

Aplicando el teorema 5.2 (condiciones KKT a funciones reales), se tiene que $\mathbf{x}^*$ es una solución óptima de la función real f bajo las restricciones del programa (MP2).
Suponiendo que $\mathbf{x}^* \notin X_{WP}^{(I)}$, esto significa que hay un $\overline{\mathbf{x}} \in X$ y $1 \leq k \leq q$ de modo que $F_k(\overline{\mathbf{x}}) \prec_{IS} F_k(\mathbf{x}^*)$, esto implica que $f(x) < f(\mathbf{x}^*)$, lo cual contradice de que f presente un valor óptimo en $\mathbf{x}^*$ En consecuencia, se tiene que $\mathbf{x}^* \in X_{WP}^{(I)}$ □

Teorema 5.8. *Dado el programa matemático (MP2), $x^* \in X$. Considerando que F es continuamente gH-diferenciable e IS-convexa en x^*. Si existen los multiplicadores de Lagrange $0 < \lambda_j \in \mathbb{R}$, $j = 1, \ldots, q$, y $0 \leqslant \mu_j \in \mathbb{R}$, $j = 1, \ldots, m$, de modo que las siguientes condiciones KKT se cumplan:*

(I) $\sum_{j=1}^{q} \lambda_j \nabla_g(F_j(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{x}^*) = 0$;

(II) $\mu_j g_j(\boldsymbol{x}^*) = 0$, *para todo* $j = 1, \ldots, m$.

entonces $\boldsymbol{x}^* \in X_P^{(I)}$ *y* $\boldsymbol{x}^* \in X_P^{(III)}$ *para el programa matemático (MP2).*

Demostración. Como F es una función multivalor intervalo, esto es
$F = (F_1, F_2, \ldots, F_q)$ donde sus componentes F_j son funciones valor intervalo ($j = 1, 2, \ldots, q$). Según consideraciones del teorema, F es continuamente gH-diferenciable en $\mathbf{x}^*$, debido a la definición (2.8), se tiene que (F_j son continuamente gH-diferenciables en $\mathbf{x}^*$ para $j = 1, \ldots, q$. y por teorema (2.4), los F_j^I y F_j^S son continuamente diferenciables en $\mathbf{x}^*$. De esto, La condición (i) sería equivalente a

$$\sum_{j=1}^{q} \lambda_j \nabla(F_j^I)(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*) = 0$$

$$= \sum_{j=1}^{q} \lambda_j \nabla(F_j^S)(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*)$$

de lo cual, sumando se obtiene

$$\sum_{j=1}^{q} \lambda_j \nabla(F_j^I)(\mathbf{x}^*) + \sum_{j=1}^{q} \lambda_j \nabla(F_j^S)(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j' \nabla g_j(\mathbf{x}^*) = 0$$

donde se considera $\mu_j' = 2\mu_j$, $j = 1, 2, \ldots, m$

A partir de esto se hace una argumentación análoga a la realizada en la demostración del teorema (5.3) y se obtiene el resultado esperado. □

Aplcación Numérica

Sea la función F multivalor intervalo dada por

$$F(x_1, x_2) = (F_1(x_1, x_2), F_2(x_1, x_2))$$

donde sus componentes son funciones de valor intervalo dadas por
$F_1(x_1, x_2) = [x_1^2 + 2x_1 + 1, x_1^2 + 2x_1 + x_2^2 - 2x_2 + 2]$
$F_2(x_1, x_2) = [x_2^2 - 2x_2 + 1, x_1^2 + 2x_1 + x_2^2 - 2x_2 + 2]$
Se tiene el siguiente programa matemático (tipo MP2)

$$\begin{aligned} min \quad & (F(x_1, x_2) = F_1(x_1, x_2), F_2(x_1, x_2)) \\ \text{sujeto a} \quad & x_1 + x_2 - 1 \leqslant 0 \\ & -x_1 - 1 \leqslant 0, \end{aligned}$$

Así se tiene

$$F_1^I(\mathbf{x}) = x_1^2 + 2x_1 + 1$$

$$F_1^S(\mathbf{x}) = x_1^2 + 2x_1 + x_2^2 - 2x_2 + 2$$

$$F_2^I(\mathbf{x}) = x_2^2 - 2x_2 + 1$$

$$F_2^S(\mathbf{x}) = x_1^2 + 2x_1 + x_2^2 - 2x_2 + 2$$

y las funciones restricción son

$$g_1(\mathbf{x}) = x_1 + x_2 - 1, \quad g_2(\mathbf{x}) = -x_1 - 1$$

Se puede ver que las funciones F_j^I, F_j^I son convexas, $j = 1, 2$. La función multivalor intervalo F es gH-diferenciable. Asimismo, las funciones restricción satisfacen las condiciones del Teorema.
Como el punto $\mathbf{x}^* = (-1, 1)$ satisface las condiciones (i) y (ii) del teorema, en consecuencia $\mathbf{x}^* = (-1, 1) \in X_P^{(I)}$ y $\mathbf{x}^* = (-1, 1) \in X_P^{(III)}$

6. Conclusiones

1. Se desarrollaron metodologías para identificar soluciones óptimas de Pareto del programa matemático (P). Se identificaron tres tipos de soluciones Pareto según la relación de orden parcial establecida sobre $\mathscr{I}$ (conjunto de todos los intervalos cerrados y acotados en $\mathbb{R}$

2. Se establecieron tres tipos de relaciones de orden parcial sobre $\mathscr{I}$, las cuales fueron las denominadas IS, CR, IW. Estas relaciones permitieron realizar las comparaciones entre los vectores con componentes de valor intervalo (elementos de $\mathscr{I}$)

3. En base a la formulación de la gH-derivada se reformularon las condiciones de Karush Khun Tucker para los programas matemáticos cuyas funciones objetivo son funciones multivalor intervalo (funciones cuyos componentes son funciones de valor intervalo)

1. Bao Y., Zao B., Bai E. (2016) Directional diferentiability of interval-valued functions, Journal of Mathematics and Computer Science 16.
2. Bazaraa M., Sherali H., Shetty C. (2006) Nonlinear programming, Theory and Algorithms, Wiley-Interscience, NY.
3. Chalco-Cano Y., Lodwick W., Rufian-Lizana A. (2013) Optimality conditions of type KKT for optimization problem with interval-valued objective function via generalized derivative, Springer Science+Business Media, New York.
4. Hosseinzade E., Hassanpour H. (2011) The Karush Kuhn Tucker Optimality conditions in interval-valued multiobjetive programming problems, J. Appl. Math. Informatics Vol. 29.
5. Ishibuchi H., Tanaka H. (1990) Multiobjective programming in optimization of the interval objective function, European Journal of Operational Research 48.
6. Moore R., Baker R., Cloud M. (2009) Introduction to interval analysis, SIAM.
7. Stefanini L., Bede B. (2009) Generalized Hukuhara Differentiability of Interval-valued Functions and Interval Differential Equations, WP-EMS Working Papers Series in Economics, Mathematics and Statistics, Vol. 3.
8. Stefanini L., Arana M. (2018) Karush-Khun-Tucker conditions for interval and fuzzy optimization in several variables under total and directional generalized differentiability article in press www.elsevier.com.
9. Wu H. (2007) The Karush Khun Tucker optimality conditions in an optimization problem with interval-valued objective function, European Journal of Operational Research 176, 2007.
10. Wu H. (2009) The Karush Khun Tucker optimality conditions in multiobjective programming problems with interval-valued objective function, European Journal of Operational Research 196, 2009.

Solución de un sistema no lineal algebráico por optimización numérica

Leopoldo Paredes Soria, Pedro Canales García
Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima, Perú,
lparedess@uni.edu.pe, pcanales@uni.edu.pe

Recibido el 02 de noviembre del 2020; aceptado el 22 de diciembre del 2020

Se analiza la solución aproximada que se obtiene al pasar de una transformación lineal a una cuadrática en el espacio de Banach lo que resulta laborioso porque se trabaja con un sistema de ecuaciones de recurrencia. Luego se procede a debilitar la función con la finalidad de generalizar el teorema de convergencia del método iterativo de Chebyshev. Se procese a analizar la nueva condición de detener los algoritmos en su ejecución, asimismo planteamos un modo de acelerar el error del teorema que se plantea en el Teorema de Convergencia. Finalmente se dan dos ejemplos de aplicación en el cual se analizará todo lo expuesto anteriormente.

Palabras Claves: Sistema no lineal, optimización, Transformación cuadrática, espacio de Banach.

We analyze the approximate solution that is obtained when passing from a linear transformation to a quadratic one in the Banach space which is laborious because we work with a system of recurrence equations. Then we proceed to weaken the function in order to generalize the theorem of convergence of Chebyshev's iterative method. It is processed to analyze the new condition to stop the algorithms in their execution, We also propose a way to accelerate the error of the theorem that arises in the Convergence Theorem. Finally, two application examples are given in which everything exposed will be analyzed. previously.

Keywords: Nonlinear system, optimization, Quadratic transformation, Banach space.

1. Introducción

La importancia del presente trabajo es mostrar como en el caso de las funciones reales cuando se desea obtener una mejor aproximación de su solución del caso lineal se pasa al caso cuadrático, en el espacio de Banach si bien la idea es la misma pero las operaciones crece de una manera exorbitante el cual pone en riesgo su utilidad y a su vez trataremos de debilitar la función para extenderlo si fuera posible, y terminaremos analizando una nueva condición de detener el algoritmo con respecto al error absoluto tradicional que se aplica, asimismo se dara una alternativa para acelerar el error del teorema, para lo cual se darán dos ejemplos de aplicación en la especialidad de la Química y de la Física.

Se está haciendo costumbre resolver sistemas de ecuaciones no lineales en la matemática computacional y en las aplicaciones, donde los modelos ha estudiar representan fenómenos que no pueden expresarse en forma lineal. Es decir, los sistemas correspondientes no todos son de la forma $Ax = b$, donde A es una matriz y x, b son vectores. Este nuevo tipo de ecuaciones algebraicas puede representarse en forma funcional compacta $f(x) = 0$, donde $f : \mathbb{R}^n \to \mathbb{R}^n$ es una función vectorial, $x = (x_1, \ldots, x_n)^T$ y $0 = (0, \cdots, 0)^T$ en $\mathbb{R}^n$. En forma desarrollada escribimos

2. Metodología

Dados las consideraciones de las ecuaciones con operadores de la forma:

$$u = T(u), \quad u \in K \tag{1}$$

Donde, K es un subespacio de V, y $T : K \to V$; V es un espacio de Banach. Las soluciones de las ecuaciones de la forma (1) son llamados punto fijo del operador T. El método más importante del análisis para la solución teórica de cada ecuación es el teorema de punto fijo de Banach.

De acuerdo a Dennis-Schnabel [1] las siguientes características deben tenerse en cuenta para desarrollar un algoritmo que resuelva un problema no lineal.

• El tamaño es un concepto que depende del procesador. Un problema se considera pequeño si tiene hasta 100 variables, mientras que si tiene entre 100 y 1000 variables se puede considerar mediano. Finalmente problemas grandes serán aquellos de más de 1000 variables. Claramente esta noción cambia a medida que cambia la tecnología. Para problemas de gran tamaño existen algoritmos especiales que explotan la estructura del problema.

• La disponibilidad de las derivadas, cuando se sabe que las funciones que intervienen en el problema son continuamente diferenciables. Sin embargo las derivadas analíticas no están disponibles o son costosas de calcular. Para eso es necesario desarrollar algoritmos que trabajan en forma eficiente ante la ausencia de derivadas.

• La eficiencia es resolver el problema costoso donde las funciones que intervienen necesitan mucho tiempo de máquina para ser evaluadas y tal vez lugar de memoria para almacenar cálculos intermedios.

También puede suceder que para la resolución del problema se necesite resolver subproblemas sencillos relacionados con él. Por lo tanto se necesitan desarrollar algoritmos que requieran pocas evaluaciones de funciones y sus derivadas, y que muestren rápida velocidad de convergencia.

• La precisión de los dígitos depende de la naturaleza del problema. En general se requieren más dígitos de los que se necesita para asegurar la convergencia del algoritmo, pero el punto es que la precisión requerida rara vez está cerca de la precisión de la máquina.
• La observación de un pobre escalamiento significa que los tamaños de las variables difieren considerablemente entre si. Si se ignora este fenómeno el comportamiento de un algoritmo para problemas no lineales se puede ver realmente afectado. Entonces, tamaño del problema, eficiencia, precisión en la solución y escalamiento del problema son características que deben ser tenidas en cuenta en el desarrollo de un algoritmo que resuelve un problema no lineal, en particular los tres problemas mencionados al principio.

2.1. Métodos Cuasi-Newton para Sistemas No Lineales

Una de las desventajas del método para sistemas no lineales es el cálculo de la matriz Jacobiana, lo que requiere una cantidad considerable de evaluaciones de funciones. En lo que sigue veremos dos estrategias, para evitar el cálculo de las derivadas.

2.2. Método de Newton con Diferencias Finitas

En el caso de una variable $f'(x)$ es aproximada por a, donde

$$a = \frac{f(x+h) - f(x)}{h}, \tag{2}$$

y h es una cantidad tal que $|f'(x) - a| \approx O(h)$.
Para el caso n-dimensional es razonable aproximar la componente $\frac{\partial f_i}{\partial x_j}$ por

$$a_{ij} = \frac{f_i(x + he_j) - f_i(x)}{h}, \quad i, j = 1, 2, \cdots, n, \tag{3}$$

donde e_j es el j-ésimo vector canónico. Esto es equivalente a aproximar la j-ésima columna de $J(x)$ por el vector

$$A_{.j} = \frac{F(x + he_j) - F(x)}{h}, \quad j = 1, 2, \cdots, n. \tag{4}$$

Se puede probar fácilmente

$$\begin{aligned} \|A_{.j} - J(x)_{.j}\|_1 &\leq C_1|h| \\ \|A - J(x)\|_1 &\leq C_1|h|. \end{aligned} \tag{5}$$

Esto nos permite construir el algoritmo del método de Newton con diferencia finita.

2.3. Método Secante para Sistemas No Lineales

Recordemos que el método de Newton aplicado a $F(x) = 0$, en cada iteración resuelve un sistema lineal

$$F'(x_c)s_c = -F(x_c), \tag{6}$$

si la solución de este sistema es x_c, el nuevo iterado es

$$x_+ = x_c + s_c. \tag{7}$$

La iteración de Newton proviene de aproximar $F(x)$ por el modelo lineal alrededor del iterado actual x_c, esto es

$$M_c(x) = F(x_c) + F'(x_c)(x_+ - x_c). \tag{8}$$

Una desventaja del método de Newton es el cálculo de $F'(x_c)$. Por otra parte, si la matriz Jacobiana es aproximada utilizando diferencias finitas, nos encontramos con el inconveniente de tener que efectuar n^2 evaluaciones de funciones.
Entonces, ¿es posible aproximar $F'(x_c)$ o su inversa, cuando éstas no están disponibles, sin efectuar evaluaciones de funciones?
Recordemos que en el método de la secante para el caso unidimensional, considerábamos los dos últimos iterados x_c, x_+ y el modelo a fín que aproxima f en un entorno de x_+,

$$m_+(x) = f(x_+) + a_+(x - x_+), \tag{9}$$

donde

$$a_+ = \frac{f(x_+) - f(x_c)}{x_+ - x_c}. \tag{10}$$

Observamos:

$$m_+(x_+) = f(x_+), \tag{11}$$

$$m_+(x_c) = f(x_c), \tag{12}$$

$$m_+(x_{++}) = 0. \tag{13}$$

El precio que se paga por esta aproximación es que se pierde la rápida convergencia cuadrática, obteniendose sólo convergencia de orden $p = \frac{1+\sqrt{5}}{2} \approx 1{,}618033989$.
Para el caso $n > 1$, se procede en forma similar. Consideremos x_c, $x_+\mathbb{R}^n$. El modelo lineal

$$M_+(x) = F(x_+) + A_+(x - x_+), \tag{14}$$

satisface $M_+(x_+) = F(x_+)$ y exigimos que A_+ sea tal

$$M_+(x_c) = F(x_c), \tag{15}$$

entonces de (14) y (15):

$$\begin{aligned} M_+(x_c) &= F(x_+) + A_+(x_c - x_+) \\ F(x_c) &= F(x_+) + A_+(x_c - x_+) \\ A_+(x_+ - x_c) &= F(x_+) - F(x_c). \end{aligned} \tag{16}$$

Dado que $x_+ - x_c = s_c$ es el paso; definiendo $y_c = F(x_+) - F(x_c)$ se tiene la ecuación de la secante

$$A_+s_c = y_c. \tag{17}$$

Este es un sistema de n ecuaciones y n^2 incógnitas, las entradas de la matriz $A_+\mathbb{R}^{n\times n}$. Por lo tanto, no se tiene una solución única.

Forma de hallar una matriz A_+

Dado que en la iteración actual no se tiene ninguna información acerca de la matriz Jacobiana o del modelo, hay que conservar la información que se tiene de las iteraciones previas. Por lo tanto el objetivo elegir A_+ tratando de minimizar el cambio en el modelo lineal satisfaciendo la ecuación de la secante.

2.4. Método de Broyden

Si de

$$A_+ = A_c + \frac{(y_c - A_c s_c)s_c^T}{s_c^T s_c}.$$

se elige v como el paso s_c^T, se tiene la actualización de Broyden

$$A_+ = A_c + \frac{(y_c - A_c s_c)v^T}{v^T s_c}. \quad (18)$$

La cual permite definir el método de Broyden que es quizá después del método de Newton el método más popular para resolver sistemas no lineales de ecuaciones algebraicas.

Actualizar la matriz indica que no se está aproximando $F'(x_+)$ ignorando $F'(x_c)$, sino que la aproximación A_c de $F'(x_c)$ está siendo corregida de modo que A_+ sea una aproximación de $F'(x_+)$.

Como $\langle s_c, n\rangle = 0$, si S es el subespacio generado por s_c, esto es $S = gen\{s_c\}$, y siendo $\langle A_+ - A_c, n\rangle = 0$, entonces $(A_+ - A_c) \in S$. Pero S es unidimensional, por lo tanto $A_+ - A_c$ debe ser una matriz de rango 1. Luego debe existir un par de vectores $u, v\mathbb{R}^n$ tal que $A_+ - A_c = uv^T$; en efecto si $u = (u_1; u_2; u_3)^T$, $v = (v_1; v_2; v_3)^T$, entonces

$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix},$$

y vemos que $rango(uv^T) = 1$.
Asi:

$$A_+ = A_c + uv^T. \quad (19)$$

Establecemos el método de Broyden.

2.5. El Método Iterativo de Kantorovich

En este estudio nos preocupamos por el problema de la aproximación de una solución única a nivel local x^* de la ecuación

$$F(x) = 0 \quad (20)$$

donde F es un operador dos veces diferenciable Fréchet $F : D \subset U \to V$, donde D es un subconjunto convexo; U y V son espacio de Banach.
El método de Newton

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, \quad n \geq 0, \quad x_0 \in D \quad (21)$$

Se ha utilizado por muchos autores (ver [2], [3], [4] y [5]) para generar una sucesión $\{x_n\}_{n\geq 0}$ convergente a x^*. En particular, las siguientes condiciones se han utilizado.

Condición A: Sea $F : D \subseteq U \to V$ es diferenciable Fréchet sobre D, $F'(x_0)^{-1} \in L(V, U)$ para algún $x_0 \in D$, donde $L(V, U)$ es el conjunto de operadores lineales acotados de V en U, y asumiendo

$$\|F'(x_0)^{-1}[F'(x) - F'(y)]\| \leq l\|x - y\|, \quad \forall x, y \in D \quad (22)$$

$$\|F'(x_0)^{-1}F(x_0)\| \leq a \quad (23)$$

y

$$2la \leq 1 \quad (24)$$

Sobre la condición A, se puede obtener el error estimado, la existencia y la unicidad de solución de las regiones, y saber si x_0 es una condición inicial contenido en dicha región, es decir, el método de Newton (21) a partir de x_0 converge a x^*. Pero a veces cuando queremos determinar si la iteración de Newton (21) a partir de x_0 converge, la condición A sería.

Condición B: Sea $F : D \subseteq U \to V$ es dos veces diferenciable Fréchet sobre D, con $F'(x) \in L(U, V)$, $F''(x) \in L(U, L(U, V))$ $(x \in D)$, $F'(x_0)^{-1}$ existe para algún $x_0 \in D$, y asumimos

$$0 < \|F'(x_0)^{-1}F(x_0)\| \leq a \quad \text{y} \quad \|F'(x_0)^{-1}F''(x_0)\| \leq b \quad (25)$$

$$\|F'(x_0)^{-1}[F'(x) - F'(x_0)]\| \leq c\|x - x_0\|, \quad c > 0 \quad (26)$$

$$\|F'(x_0)^{-1}[F''(x) - F''(x_0)]\| \leq d\|x - x_0\|, \quad \forall x \in D \quad (27)$$

y

$$2La \leq 1 \quad (28)$$

donde sea

$$L = \text{máx}\{c, b + 2ad\} \quad (29)$$

o, si la función

$$f(t) = t^3 - 2bt^2 - (2d - b^2)t + 2d(b + ad) \quad (30)$$

tiene dos raices positivos r_1 y r_2 tal que:

$$[b, b + 2ad] \subseteq [r_1, r_2] \quad (31)$$

entonces $L \geq c$ y

$$L \in [b, b + 2ad] \quad (32)$$

2.6. Método Iterativo de Chebyshev

Sean X y Y espacios de Banach y $F : \Omega \subseteq X \longrightarrow Y$ un operador no lineal, el cual es diferenciable Fréchet sobre un dominio convexo abierto Ω. Supongamos que $F'(x_0)^{-1} \in L(Y, X)$ existe para algún $x_0 \in \Omega$, donde $L(Y, X)$ es el conjunto de operadores lineales acotados de Y sobre X.
El proceso iterativo más famoso que aproxima a la solución $x^* \in \Omega$ de la ecuación:

$$F(x) = 0, \quad (33)$$

es el método de Newton:

$$x_{n+1} = x_n - F'(x_n)^{-1}F(x_n), \quad n \geq 0. \quad (34)$$

El resultado básico concerniente a la convergencia del método de Newtón de la existencia y unicidad de soluciones, y la estimación del error son dados por el Teorema de Kantorovich (ver [5], [6], [7] y [8]).

Por otra parte, Candela y Marquina [9], [10] construyen un conjunto de sucesiones que satisfacen algunas relaciones recurrentes, con la cual prueba que la sucesión (34)

está bien definida y converge para una solución x^* de (33) si

$$a = k\beta\eta \in \left[0, \frac{1}{2}\right\rangle, \tag{35}$$

donde

$$\begin{aligned} \|F'(x_0)^{-1}\| &\leq \beta, \\ \|F'(x_0)^{-1}F(x_0)\| &\leq \eta, \quad \text{y} \\ \|F'(x) - F'(y)\| &\leq k\|x - y\|. \end{aligned} \tag{36}$$

Bajo la misma condición (35) como el método de Newton y con el mismo costo operacional, construimos un nuevo método iterativo para resolver (33). Esta iteración está definida por:

$$\begin{aligned} &\Gamma_n = F'(x_n)^{-1},\ T(x_n) = \frac{1}{2}\Gamma_n A\Gamma_n F(x_n), \\ &x_{n+1} = x_n - [I + T(x_n)]\Gamma_n F(x_n),\ n \geq 0 \end{aligned} \tag{37}$$

Además, observamos que si la segunda derivada Fréchet es reemplazada en el método Chebyshev [1], [11] por el operador bilineal A, la ecuación (37) es también obtenido, y consecuentemente el proceso iterativo es llamado método iterativo de Chebyshev. Es decir, se debe a la aceleración convexa de los métodos aplicada a las ecuaciones cuadráticas, en la cual la segunda derivada Fréchet es reemplazado por el operador bilinel A.

3. Aportes

3.1. Deducción del Método de Segundo Orden

Por el desarrollo de Taylor hasta el segundo orden, se tiene:

$$F(x_{n+1}) = F(x_n) + F'(x_n)h + F''(x_n)\frac{h^2}{2}$$

Se cumple que $F(x_{n+1}) = 0$ y $F''(x_n) = A$, logrando:

$$-F(x_n) - A\frac{h^2}{2} = F'(x_n)h$$

Sabemos que $h = -F'(x_n)^{-1}F(x_n) = -\Gamma_n F(x_n)$ al reemplazar:

$$-F'(x_n)^{-1}F(x_n) - \frac{1}{2}F'(x_n)^{-1}A(-F'(x_n)^{-1}F(x_n))^2 = h$$

$$-\Gamma_n F(x_n) - \frac{1}{2}\Gamma_n A\Gamma_n F(x_n)\Gamma_n F(x_n) = h$$

$$-[I + T(x_n)]\Gamma_n F(x_n) = x_{n+1} - x_n$$

Finalmente

$$x_{n+1} = x_n - [I + T(x_n)]\Gamma_n F(x_n),\ n \geq 0$$

3.2. Debilitamiento de la función F

Se busca debilitar la función F como una experimentación y poder ver si es posible obter un nuevo enfoque de los conceptos dados en el capítulo anterior.
En el teorema de convergencia cuando debilitamos a F que es una función convexa con una cuasi-convexa tenemos:

$$F(\lambda x + (1-\lambda)y) \leq \text{máx}\{F(x), F(y)\},\ \forall\lambda \in [0,1].$$

Puede ocurrir los siguientes casos:
- Como toda función convexa es cuasi-convexa. En este caso diremos que el teorema de convergencia se cumple.
- Existen funciones no convexas y que son cuasi-convexas. En este caso diremos que el teorema de convergencia no se cumple, porque F no tiene derivada Fréchet en los puntos picos.
- Además hay funciones discontinuas que son cuasi-convexas. En este caso diremos que el teorema de convergencia no se cumple, porque F no es continua y no tiene derivada Fréchet en dicho punto.

Por consiguiente la convexidad de la función F está implicito en el Teorema de Convergencia por ser de suma importancia.

3.3. Nueva Forma de Parada del Algoritmo

La construcción de las sucesiones $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ y $\{d_n\}$ que son acotadas, el cual permite la convergencia del método iterativo de Chebyshev donde analizaremos y recomendaremos su utilización a partir de adelante. Las variantes que usamos se mencionan a continuación.
- Usando b_n como Parada vs Error Absoluto Si bien se puede usar la sucesión $\{b_n\}$ como parada en el algoritmo ya que garantiza

$$\|T(x_n)\| \leq b_n.$$

Notamos que el error absoluto es más efectivo que el error del método.
- Usando b_n como Parada vs Error Si bien se puede usar la sucesión $\{b_n\}$ como parada en el algoritmo ya que garantiza

$$\|T(x_n)\| \leq b_n.$$

- Usando $d_n\eta$ como Parada vs el Error Absoluto Si bien se puede usar $d_n\eta$ como parada en el algoritmo ya que garantiza

$$\|x_{n+1} - x_n\| \leq d_n\eta.$$

- Usando $d_n\eta$ como Parada vs Error Si bien se puede usar $d_n\eta$ como parada en el algoritmo ya que garantiza

$$\|x_{n+1} - x_n\| \leq d_n\eta.$$

En todos los casos analizados se tendrá los mismos valores de las sucesiones $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ y $\{d_n\}$ que son.

A pesar que para este ejemplo los errores del método fueron tan eficientes se recomienda usarlos porque propios del calculos de las sucesiones obtenidas y garantizan la convergencia del método.

3.4. Mejoramiento del Error del Teorema de Convergencia

En este proceso agregaremos en el programa lo siguientes pasos:

$$\beta = a_n\beta; \quad \eta = c_n\eta;$$

Con el cual se mejora el error del método, su utilidad se garantiza porque mientras que β_n crece debido que $a_n > 1$, los valores de η_n decrece, es lo que permite mejorar el error de método y lo cual se muestra en los ejemplos de aplicación que damos a continuación.

4. Aplicaciones

Se presentan aplicaciones tomando en cuenta la nueva condición para detener el algoritmo del teorema de convergencia del método iterativo de Chebyshev que se planteo en el capítulo anterior el cual difiere del error absoluto tradicional que se aplica y por ende merece analizar su utilidad.
Además se busca mostrar posibles trabajos de investigación multidisciplinarios con investigaciones del área de Química.
La resolución de problemas es de suma importancia para el avance de las matemc'aticas así como su comprensión y aprendizaje. En el antegrado se resuelven usando los métodos de Jacobi, Gauss Seidel y SOR para sistemas de ecuaciones no lineales.
Asimismo cabe indicar que cuando llegue al equilibrio la concentración a la salida del reactor, las constantes de equilibrio son fijos.

4.1. Reacciones en fase gaseosa

En un reactor se efectua las siguientes reacciones en fase gaseosa:

$$\begin{aligned} A + B &\rightleftarrows C + D \\ A + C &\rightleftarrows 2E \end{aligned}$$

A la temperatura de la reacción, las constantes de equilibrio son $kp_1 = 2{,}6$ y $kp_2 = 3{,}1$. Las composiciones iniciales son 2 $\frac{\text{mol}}{\text{l de A}}$ y 1 $\frac{\text{mol}}{\text{l de B}}$. Determine la concentración a la salida del reactor, asumiendo que se alcanza el equilibrio usando sistemas de ecuaciones no lineales, usando los métodos de Newton Kantorovich y Chebyshev.
La ley de acción de masa que establece que para una reacción reversible en equilibrio y a una temperatura constante, una relación determinada de concentraciones de reactivos y productos cuando estos llegan al equilibrio tiene un valor constante que es llamado constante de equilibrio.
Así, para una temperatura constante se cumple que:

$$K_c = \frac{[C]^c[D]^d}{[A]^a[B]^b},$$

donde K es la constante de equilibrio y $[X]$ indica la concentración de la especie X en el equilibrio.
Para un equilibrio homogéneo gaseoso, se tomará en cuenta las presiones parciales de las especies reactivas. Así, para nuestra ecuación tenemos:

$$K_p = \frac{[PC]^c[PD]^d}{[PA]^a[PB]^b},$$

donde $[PX]$ indica la presión parcial de la especie X.
La relación entre ambas constantes de equilibrio está dada por:

$$K_p = K_c * (RT)^{(c+d-a-b)};$$

donde R es la constante universal de los gases ideales y T es la temperatura.
Las variables a usar son:
x : Cambio de concentración en la primera reacción.
y : Cambio de concentración en la segunda reacción.
Las cantidades de las concentraciones en las dos reacciones en el equilibrio son:

Concentración	Masas
n_A	$2 - x - y$
n_B	$1 - x$
n_C	$x - y$
n_D	x
n_E	$2y$

El cambio en el número de moles del gas en las dos ecuaciones químicas balanceadas es igual a 0, entonces se da que $K_p = K_c$.
Por la ley de acción de masas se tiene:

$$2{,}6 = K_{p1} = K_{c1} = \frac{[C][D]}{[A][B]} = \frac{(x-y)x}{(2-x-y)(1-x)}$$

$$3{,}1 = K_{p2} = K_{c2} = \frac{[E]^2}{[A][C]} = \frac{(2y)^2}{(2-x-y)(x-y)}$$

Planteando el siguiente sistema de ecuaciones no lineales de $F : \mathbb{R}^2 \to \mathbb{R}^2$ con:

$$f_1(x,y) = 1{,}6x^2 + 3{,}6xy - 7{,}8x - 2{,}6y + 5{,}2 = 0,$$

$$f_2(x,y) = 3{,}1x^2 + 0{,}9y^2 - 6{,}2x + 6{,}2y = 0.$$

Entonces

$$F(x,y) = (1{,}6x^2 + 3{,}6xy - 7{,}8x - 2{,}6y + 5{,}2,$$

$$3{,}1x^2 + 0{,}9y^2 - 6{,}2x + 6{,}2y)$$

4.1.1. Convergencia usando el Método de Newton-Kantorovich.

De la condición B, tenemos.
Sea el $D = \langle -0{,}34; 1{,}94\rangle \times \langle -0{,}74; 1{,}54\rangle$ con $\overline{x}_0 = (0{,}8; 0{,}4) \in \Omega \subseteq D$.
El Jacobiano es:

$$F'(\overline{x}) = \begin{bmatrix} 3{,}2x + 3{,}6y - 7{,}8 & 3{,}6x - 2{,}6 \\ 6{,}2x - 6{,}2 & 1{,}8y + 6{,}2 \end{bmatrix}$$

La inversa de la matriz Jacobiana es:

$$F'(\overline{x})^{-1} = P \begin{bmatrix} 0{,}9y + 3{,}1 & 1{,}3 - 1{,}8x \\ 3{,}1(1-x) & 1{,}6x + 1{,}8y - 3{,}9 \end{bmatrix}$$

con

$$P = [2(0{,}9y + 3{,}1)(1{,}6x + 1{,}8y - 3{,}9) - 6{,}2(x-1)(1{,}8x - 1{,}3)]^{-1}.$$

El valor de $\|F'(\overline{x}_0)^{-1}\|_\infty = 0{,}277469478357381 = \beta$
Los valores son:

$$\|F'(\overline{x}_0)^{-1}F(\overline{x}_0)\|_\infty = \left\| \begin{array}{c} -0{,}029399432729066 \\ -0{,}056135158465902 \end{array} \right\|_\infty = 0{,}056135158465902$$

$$\Rightarrow a = 0{,}056135158465902.$$

$$F''(\overline{x}_0) = \left[\begin{array}{cc} 3{,}2 & 3{,}6 \\ 3{,}6 & 0 \\ \hline 6{,}2 & 0 \\ 0 & 1{,}8 \end{array}\right]$$

Luego:

$$\|F'(\overline{x}_0)^{-1}F''(\overline{x}_0)\|_\infty = 1{,}774571463805648 = b.$$

Donde:

$$\|F'(\overline{x}_0)^{-1}[F'(\overline{x}) - F'(\overline{x}_0)]\|_\infty =$$

$$\left\| F'(\overline{x}_0)^{-1} \left[\begin{array}{ll} 3{,}2(x-0{,}8) + 3{,}6(y-0{,}4) & 3{,}6(x-0{,}8) \\ 6{,}2(x-0{,}8) & 1{,}8(y-0{,}4) \end{array} \right] \right\|_\infty$$

$$\leq \left\| F'(\overline{x}_0)^{-1} \left[\begin{array}{cc} 3{,}2 & 3{,}6 \\ 3{,}6 & 0 \\ \hline 6{,}2 & 0 \\ 0 & 1{,}8 \end{array}\right] \right\|_\infty \|\overline{x} - \overline{x}_0\|_\infty$$

$$\leq 1{,}77457146380568\|\overline{x} - \overline{x}_0\|_\infty$$

$$\Rightarrow c = 1{,}77457146380568.$$

Dado que $F''(\overline{x})$ es constante, entonces $d = 0$. Además:

$$L = \text{máx}\{c, b + 2ad\} = \text{máx}\{c, b\} = 1{,}774571463805648,$$

con

$$2La = 0{,}199231700659596 < 1.$$

De

$$\begin{aligned} f(t) &= t^3 - 2bt^2 - (2d - b^2)t + 2d(b + ad) \\ &= t^3 - 2bt^2 + b^2t \\ &= t(t-b)^2 \end{aligned}$$

$$\Rightarrow t = b = 1{,}774571463805648 = r_1 = r_2.$$

Se cumple:

$$[b, b + 2ad] = [b, b] \subseteq [r_1, r_2] \Rightarrow [b, b] \subseteq [r_1, r_1].$$

Entonces $L \geq c$ y $L \in [b, b + 2ad] \Rightarrow L \in [b, b]$.

Luego, tenemos:

$$p(t) = 0{,}887285731902824t^2 - t + 0{,}056135158465902.$$

Luego, la sucesión nos da una raíz, con $t_0 = 0$:

$$t_{n+1} = t_n - \frac{0{,}887285731902824t_n^2 - t_n + 0{,}056135158465902}{1{,}774571463805648t_n - 1}$$

n	t_n	E	$p(t_n)$
0	0		0,056135158
1	0,056135158	0,056135158	0,002795975
2	0,059240472	0,003105314	0,000008556
3	0,059250033	9,561212107	$8{,}1113 \times 10^{-11}$
4	0,059250034	$9{,}0643 \times 10^{-11}$	0,000000000

donde $2La = 0{,}199231700659596 < 1$.

Entonces

$$p(0{,}05925003464793) = 0 \Rightarrow r_1 = 0{,}059250034064793.$$

Análogamente

n	t_n	E	$p(t_n)$
0	1		−0,056579109
1	1,073045693	0,073045693	0,004734267
2	1,06780980	0,005235885	0,000024324
3	1,067782628	0,000027181	$6{,}55537 \times 10^{-10}$
4	1,067782627	$7{,}32560 \times 10^{-10}$	0,000000000

$$\Rightarrow r_2 = 1{,}06778262750574$$

Entonces, el método de Newton-Kantorovich está bien definido, la solución $\overline{x}^* \subset \Omega$ y es única en el conjunto $\{\overline{x}/\|\overline{x} - \overline{x}_0\| \leq r_1\} \cap \Omega$, donde $r_1 = 0{,}059250034064793$.

La tabla del método de Newton-Kantorovich es:

k	x_k	y_k	Error
0	0,8	0,4	
1	0,829399432729	0,4561351584659	0,056135158465
2	0,831437055478	0,4556565669837	0,002037622749
3	0,831437753043	0,4556548080497	0,000001758934
4	0,831437753042	0,4556548080488	0,000000000001

4.2. Convergencia del Método Iterativo de Chebyshev.

De las condiciones del teorema.
Sea el $\Omega = D = \langle -0{,}34; 1{,}94\rangle \times \langle -0{,}74; 1{,}54\rangle$ con $\overline{x}_0 = (0{,}8; 0{,}4) \in \Omega$.
El Jacobiano es:

$$F'(x,y) = JF(x,y) = \left[\begin{array}{ll} 3{,}2x + 3{,}6y - 7{,}8 & 3{,}6x - 2{,}6 \\ 6{,}2x - 6{,}2 & 1{,}8y + 6{,}2 \end{array} \right]$$

La inversa de la matriz Jacobiana es:

$$JF(\overline{x})^{-1} = P \left[\begin{array}{ll} 0{,}9y + 3{,}1 & 1{,}3 - 1{,}8x \\ 3{,}1(1-x) & 1{,}6x + 1{,}8y - 3{,}9 \end{array} \right]$$

con

$$P = (2(0{,}9y + 3{,}1)(1{,}6x + 1{,}8y - 3{,}9) - 6{,}2(x-1)(1{,}8x - 1{,}3))^{-1}.$$

El valor de $\|J(\overline{x}_0)^{-1}\|_\infty = 0{,}277469478357381 = \beta$
Los valores son:

$$\|JF(\overline{x}_0)^{-1}F(\overline{x}_0)\|_\infty = \left\| \begin{array}{c} -0{,}029399432729066 \\ -0{,}056135158465902 \end{array} \right\|_\infty = 0{,}056135158465902$$

$$\Rightarrow \eta = 0{,}056135158465902.$$

Luego:

$$F'(x,y) - F'(u,v) = \begin{bmatrix} 3{,}2x + 3{,}6y - 7{,}8 & 3{,}6x - 2{,}6 \\ 6{,}2x - 6{,}2 & 1{,}8y + 6{,}2 \end{bmatrix} - \begin{bmatrix} 3{,}2u + 3{,}6v - 7{,}8 & 3{,}6u - 2{,}6 \\ 6{,}2u - 6{,}2 & 1{,}8v + 6{,}2 \end{bmatrix}$$

$$= \begin{bmatrix} 3{,}2(x-u) + 3{,}6(y-v) & 3{,}6(x-u) \\ 6{,}2(x-u) & 1{,}8(y-v) \end{bmatrix} = \left[\begin{array}{cc} 3{,}2 & 3{,}6 \\ 3{,}6 & 0 \\ \hline 6{,}2 & 0 \\ 0 & 1{,}8 \end{array}\right] \begin{bmatrix} x-u \\ y-v \end{bmatrix}$$

Tomando norma m.a.m. tenemos:

$$\|F'(x,y) - F'(u,v)\|_\infty \leq \left\| \begin{array}{cc} 3{,}2 & 3{,}6 \\ 3{,}6 & 0 \\ \hline 6{,}2 & 0 \\ 0 & 1{,}8 \end{array} \right\|_\infty \|(x,y)-(u,v)\|_\infty.$$

$$\Rightarrow k = 6{,}8$$

Si $a = k\beta\eta = 0{,}105915393331891 < 0{,}5$ y $\alpha = 3{,}4 \in \langle 0, 72{,}66851449131069\rangle$.

Entonces $\{x_n\}$ está bien definido, $x_n \in B(x_0, r\eta)$, $\forall n \geq 0$ y converge para $x^* \in \overline{B}(x_0, r\eta)$.

De la relación recurrente, se requiere

$$b = \alpha\beta\eta = 0{,}05295769666594540$$

donde $a_0 = 1$, $b_0 = \frac{b}{2} = 0{,}02647884833297270$, $c_0 = 1$ y $d_0 = 1 + \frac{b}{2} = 1{,}02647884833297270$.

La tabla es:

k	a_k	b_k	c_k	d_k
0	1,000000000	0,0264788483	1,000000000	1,026478848
1	1,121981756	0,0027425555	0,092314631	0,092567809
2	1,134461179	0,0000231540	0,000770794	0,000770812
3	1,134566261	0,0000000016	0,000000054	0,000000053
4	1,134566268	1×10^{-17}	$2{,}6 \times 10^{-17}$	$2{,}6 \times 10^{-17}$

el resultado del método iterativo de Chebyshev es:

k	x_k	y_k	*Error*	*Error Teor.*
0	0,800000000	0,400000000		
1	0,830638409	0,456039843	0,056039843	0,005196309
2	0,831437731	0,455655015	0,000799322	0,000043270
3	0,831437753	0,455654808	0,000000206	0,000000003
4	0,831437753	0,455654808	$3{,}16 \times 10^{-16}$	1×10^{-17}

Se observa que cuando en el algoritmo agregamos $\beta = a_n\beta$ y $\eta = c_n\eta$ y realizamos los calculos lo que mejora es el error del teorema, como mostramos a continuación en las tablas siguientes:

k	a_k	b_k	c_k	d_k
0	1,0000	0,0264788	1,000000	1,02647
1	1,1219	0,0027425	0,092314	0,09256
2	1,1344	0,0000231	0,000771	0,00077
3	1,1346	$1{,}6 \times 10^{-9}$	$5{,}35476604 \times 10^{-9}$	$5{,}35476605 \times 10^{-9}$

el resultado del método iterativo de Chebyshev es:

k	x_k	y_k	*Error*	*Error Teorema*
0	0,80000000	0,400000000		
1	0,83063840	0,456039842	0,0560398	0,000479695
2	0,83143773	0,455655014	0,0007993	0,000000003
3	0,83143775	0,455654808	0,0000002	0,000000000

4.3. Reflexión difusa en un medio paralelo

La teoria de la transferencia de radiación discutida hasta ahora utiliza la formulación local, ya que se emplean cantidades definidas solo localmente. En 1943 Ambarzumian formuló un principio de invariancia para la ley de reflexión difusa:

• La ley de la reflexión difusa de un medio paralelo plano homogéneo infinitamente profundo es invariante con respecto a la adición (o sustracción) de capas de espesor óptico finito arbitrario hacia (o desde) el medio.

Para la ley del oscurecimiento (es debido a la absorción y difusión de la luz de la estrella [5], el profesor Chandrasekhar modificó el principio de la siguiente manera:

• La distribución emergente de un medio paralelo plano semi-infinito es invariante a la suma (o resta) de capas de espesor óptico arbitrario hacia (o desde) el medio.

Consideremos el espacio $X = C[0,1]$ de las funciones continuas en $[0,1]$ y dotado de la norma del máximo $\|x\| = \max_{t\in[0,1]} |x(t)|$, $x \in X$. Se desea encontrar una función $x \in X$ que obedece a la ecuación integral no lineal.

$$x(s) = 1 + \lambda x(s) \int_0^1 \frac{s}{s+t} x(t)dt, \ s \in [0,1]. \tag{38}$$

La resolución de (38) es equivalente a resolver (20), siendo $F : X \to X$ y tal que:

$$F(x)(s) = x(s) - 1 - \lambda x(s) \int_0^1 \frac{s}{s+t} x(t)dt, \ s \in [0,1]. \tag{39}$$

Estudiamos el caso particular $\lambda = \frac{1}{4}$, por conveniencia. Veamos la existencia y unicidad de la solución de esta ecuación.
Para aplicar el teorema de Kantorovich, necesitamos partir de una función inicial adecuada. Dado (38) se deduce que $x(0) = 1$, una elección razonable parece ser $x_0(s) = 1$, $\forall s \in [0,1]$. Además:

$$\begin{aligned} F'(x)y(s) =& y(s) - \frac{1}{4}x(s) \int_0^1 \frac{s}{s+t} y(t)dt \\ & - \frac{1}{4}y(s) \int_0^1 \frac{s}{s+t} x(t)dt, \ y \in [0,1]. \end{aligned} \tag{40}$$

$$\begin{aligned} F''(x)yz(s) =& -\frac{1}{4}z(s) \int_0^1 \frac{s}{s+t} y(t)dt - \\ & \frac{1}{4}y(s) \int_0^1 \frac{s}{s+t} z(t)dt, \ z \in [0,1]. \end{aligned} \tag{41}$$

Podemos considerar el subconjunto

$$\Omega \subseteq D = \underbrace{\langle -0{,}5, 2{,}5\rangle \times \langle -0{,}5, 2{,}5\rangle \cdots \langle -0{,}5, 2{,}5\rangle}_{8 \text{ términos}},$$

porque vemos que es conveniente, que aparece en el enunciado del teorema de Kantorovich como el propio espacio X, que es un conjunto convexo. Procedemos a calcular las constantes a, b, c y d.

En primer lugar,

$$\begin{aligned} F(x_0)(s) =& x_0(s) - 1 - \frac{1}{4}x_0(s)\int_0^1 \frac{s}{s+t}x_0(t)dt \\ =& -\frac{s}{4}\int_0^1 \frac{1}{s+t}dt \\ =& -\frac{s}{4}\ln\left(\frac{s+1}{s}\right) \end{aligned} \tag{42}$$

tomando la norma del máximo m.a.m. tenemos.

$$\|F(x_0)\| = \frac{\ln(2)}{4} = 0{,}173286795139. \tag{43}$$

Por el lema de Banach sobre inversión de operadores nos permite encontrar una cota para $\Gamma_0 = F'(x_0)^{-1}$. Como lo mostraremos: Dada una función cualquiera $y \in X$, se tiene que:

$$\begin{aligned} \|[I - F'(x_0)]y\| &= \max_{s\in[0,1]} |y(s) - F'(x_0)y(s)| \\ &= \frac{1}{4}\max_{s\in[0,1]}\left|\int_0^1 \frac{s}{s+t}y(t)dt \right. \\ &\left. +y(s)\int_0^1 \frac{s}{s+t}dt\right| \\ &\leq \frac{\ln(2)}{2}\|y\|, \end{aligned} \tag{44}$$

luego

$$\|I - F'(x_0)\| \leq \frac{\ln(2)}{2} = 0{,}346573590279 \ < \ 1, \tag{45}$$

con lo que se garantiza que existe $\Gamma_0 = F'(x_0)^{-1}$ y además

$$\begin{aligned} \|\Gamma_0\| &\leq \frac{1}{1 - \|I - F'(x_0)\|} \leq \frac{1}{1 - 0{,}346573590279} \\ &= 1{,}530394219032. \end{aligned} \tag{46}$$

En consecuencia,

$$\|\Gamma_0 F(x_0)\| \leq \frac{\frac{\ln(2)}{4}}{1 - \frac{\ln(2)}{2}} = 0{,}265197109516 = a, \tag{47}$$

y

$$\begin{aligned} \|\Gamma_0 F''(x_0)yz\| &\leq \|\Gamma_0\|\|F''(x_0)\|\|y\|\|z\| \\ &\leq \frac{\frac{\ln(2)}{2}}{1 - \frac{\ln(2)}{2}}\|y\|\|z\| \end{aligned} \tag{48}$$

$$\Rightarrow \|\Gamma_0 F''(x_0)\| \leq 2a = 0{,}530394219033 = b, \tag{49}$$

luego

$$\begin{aligned} \|\Gamma_0[F'(x) - F'(x_0)]y\| \leq& \|\Gamma_0\|\|F'(x) - F'(x_0)\|\|y\| \\ &\leq \frac{\frac{\ln(2)}{2}}{1 - \frac{\ln(2)}{2}}\|y\| \end{aligned} \tag{50}$$

$$\|\Gamma_0[F'(x) - F'(x_0)]\| \leq b = 0{,}530394219033 = c. \tag{51}$$

No olvidemos que la $F''(x)$ es constante, entonces en forma parecida deducimos que

$$\begin{aligned} [F''(x) - F''(x_0)]yz = &-\frac{y}{4}\int_0^1 \frac{s}{s+t}z(t)dt \\ &-\frac{z}{4}\int_0^1 \frac{s}{s+t}y(t)dt \\ +\frac{y}{4}\int_0^1 \frac{s}{s+t}z(t)dt &+ \frac{z}{4}\int_0^1 \frac{s}{s+t}y(t)dt = 0yz. \end{aligned} \tag{52}$$

Completando

$$\|\Gamma_0[F''(x) - F''(x_0)]yz\| \leq \|\Gamma_0\|\|[F''(x) - F''(x_0)]\|\|yz\| \tag{53}$$

$$\Rightarrow \|\Gamma_0[F''(x) - F''(x_0)]\| \leq 0 = d. \tag{54}$$

Luego

$$L = \max\{c, b + 2ad\} = \max\{c, b\} = 0{,}530394219033, \tag{55}$$

Como $a, L > 0$. Tenemos:

$$p(t) = 0{,}265197109517251t^2 - t + 0{,}265197109517251.$$

Luego, la sucesión nos da una raíz, con $t_0 = 0$:

$$t_{n+1} = t_n - \frac{0{,}265197109517251t_n^2 - t_n + 0{,}265197109517251}{0{,}530394219034502t_n - 1}$$

n	t_n	E	$p(t_n)$
0	0		0,2651971095
1	0,2651971095	0,26519710952	0,018651182
2	0,2869011621	0,02170405257	0,000124925
3	0,2870485093	0,00014734725	0,000000006
4	0,2870485161	0,00000000679	0,0

donde $2La = 0{,}281318027584 < 1$.
Entonces

$$p(0{,}287048516133531) = 0 \ \Rightarrow \ r_1 = 0{,}287048516133531,$$

la otra raíz se puede obtener por:

$$r_2 = \frac{1 + \sqrt{1 - 2La}}{L} = 3{,}483731647428. \tag{56}$$

Entonces, la ecuación de Chandrasekhar (38) está bien definido, en el sentido donde la solución $x^* \subset \Omega$ y es única en el conjunto $\{x/\|x - x_0\| \leq r_1\} \cap \Omega$, donde $r_1 = 0{,}287048516133$.
Ahora vamos encontrar la ecuación de sistemas no lineales de (38), para lo cual utilizaremos la fórmula de Gauss-Legendre que permitirá obtener el operador bilineal A, trabajamos sobre:

$$\int_0^1 f(t)dt = \int_0^1 \frac{x(s)}{4}\frac{s}{s+t}x(t)dt, \tag{57}$$

haciendo $b = 1$ y $a = 0$, tenemos:

$$t = \frac{1-0}{2}u + \frac{1+0}{2}, u \in [-1, 1], \tag{58}$$

con

$$g(u) = f\left(\frac{u}{2} + \frac{1}{2}\right). \tag{59}$$

Reemplazando, se cumple:

$$\int_0^1 f(t)dt = \frac{1}{2}\int_{-1}^1 g(u)du = \frac{1}{2}\sum_{j=1}^m w_j g(u_j), \tag{60}$$

donde t_j y w_j son los nodos y pesos conocidos. Consideraremos $m = 8$ para obtener el cuadro siguiente:

j	u_j	t_j	w_j
1	−0,960289857	0,0198550715	0,101228536
2	−0,796666478	0,1016667610	0,222381034
3	−0,525532410	0,2372337950	0,313706646
4	−0,183434642	0,4082826790	0,362683783
5	0,183434642	0,5917173210	0,362683783
6	0,525532410	0,7627662050	0,313706646
7	0,796666478	0,8983332390	0,222381034
8	0,960289857	0,9801449285	0,101228536

Denotamos x_i a las aproximaciones $x(t_i)$, $i = 1, \cdots, 8$, logrando obtener el sistema de ecuaciones no lineal siguiente:

$$x_i = 1 + \frac{1}{8}x_i\sum_{j=1}^8 w_j\frac{t_i}{t_i + t_j}x_j,\ i = 1, \cdots, 8. \tag{61}$$

Haciendo $a_{ij} = \frac{w_j t_i}{8(t_i+t_j)}$, tenemos.

$$x_i = 1 + x_i\sum_{j=1}^8 a_{ij}x_j,\ i = 1, \cdots, 8. \tag{62}$$

El sistema de ecuaciones no lineales es:

$$F(x_i) = 1 + x_i\sum_{j=1}^8 a_{ij}x_j - x_i,\ i = 1, \cdots, 8, \tag{63}$$

donde A es un operador bilineal de orden 64×8, el cual no es tan facil de ser representado.

La tabla del método de Newton-Kantorovich es:

k	0	1	2	3	4
x_{1_k}	1,0	1,021632	1,021720	1,021720	1,021720
x_{2_k}	1,0	1,072393	1,073186	1,073186	1,073186
x_{3_k}	1,0	1,123421	1,125724	1,125725	1,125725
x_{4_k}	1,0	1,165547	1,169750	1,169753	1,169753
x_{5_k}	1,0	1,197036	1,203067	1,203079	1,203072
x_{6_k}	1,0	1,218969	1,226484	1,226491	1,226491
x_{7_k}	1,0	1,232961	1,241516	1,241525	1,241525
x_{8_k}	1,0	1,240309	1,249439	1,249449	1,249449
$Error$		0,265197	0,021704	0,000147	0,000000007

Veamos la convergencia del método iterativo de Chebyshev.
De las condiciones del teorema.
Sea el $\Omega = D = \langle -0{,}5; 2{,}5\rangle \times \langle -0{,}5; 2{,}5\rangle \cdots \langle -0{,}5; 2{,}5\rangle$, tomando convenientemente $\overline{x}_0 = (1; 1; 1; 1; 1; 1; 1; 1)^T \in \Omega$.
El Jacobiano evaluado en $\overline{x}_0$ es:

$$\|\Gamma_0\| \leq \frac{1}{1 - \frac{\ln(2)}{2}} = 1{,}53039421903450235 = \beta, \tag{64}$$

luego

$$\|\Gamma_0 F(\overline{x}_0)\| \leq \frac{\frac{\ln(2)}{4}}{1 - \frac{\ln(2)}{2}} = 0{,}26519710951725117 = \eta, \tag{65}$$

sabemos que:

$$\|F'(\overline{x}) - F'(\overline{u})\| \leq \frac{\ln(2)}{2}\|\overline{x} - \overline{u}\| \tag{66}$$

$$\Rightarrow k = \frac{\ln(2)}{2} = 0{,}34657359027997264. \tag{67}$$

Si $a = k\beta\eta = 0{,}14065901379260984 < 0{,}5$ y $\alpha = 0{,}09096495075 \in \langle 0, 2{,}410077214714\rangle$ obtenido de (63).

Entonces $\{\overline{x}_n\}$ está bien definido, $\overline{x}_n \in B(\overline{x}_0, r\eta)$, $\forall n \geq 0$ y converge para $\overline{x}^* \in \overline{B}(\overline{x}_0, r\eta)$.

De la relación recurrente, se requiere:

$$b = \alpha\beta\eta = 0{,}03691868226846743$$

donde

$$a_0 = 1,\ b_0 = \frac{b}{2} = 0{,}01845934113423372,\ c_0 = 1$$

y $d_0 = 1 + \frac{b}{2} = 1{,}01845934113423375$.

La tabla es:

k	a_k	b_k	c_k	d_k
0	1,00000000	0,0184593411	1,00000000	1,01845934
1	1,16720911	0,0022988141	0,10669375	0,10693902
2	1,18806812	0,0000264311	0,00120519	0,00120523
3	1,18830746	0,0000000034	0,00000015	0,00000015
4	1,18830749	5×10^{-18}	0,0	0,0

el resultado del método iterativo de Chebyshev es:

k	0	1	2	3	4
x_{1_k}	1,0	1,0224	1,0217	1,0217	1,0217
x_{2_k}	1,0	1,0743	1,0732	1,0732	1,0732
x_{3_k}	1,0	1,1261	1,1257	1,1257	1,1257
x_{4_k}	1,0	1,1700	1,1698	1,1698	1,1698
x_{5_k}	1,0	1,1993	1,2031	1,2031	1,2031
x_{6_k}	1,0	1,2224	1,2265	1,2265	1,2265
x_{7_k}	1,0	1,2362	1,2415	1,2415	1,2415
x_{8_k}	1,0	1,2432	1,2494	1,2494	1,2494
$E - T$		0,02836	0,00032	$4{,}06 \times 10^{-8}$	0,0

Se observa que cuando en el algoritmo agregamos $\beta = a_n\beta$ y $\eta = c_n\eta$ y realizamos los calculos lo que mejora es el error del teorema, como mostramos a continuación en las tablas siguientes:

k	a_k	b_k	c_k	d_k
0	1,000000	0,018459341	1,0000000	1,018459
1	1,167209	0,002298814	0,1066938	0,106939
2	1,188068	0,000026431	0,0012052	0,001205
3	1,188307	0,000000003	0,0000002	0,000000

el resultado del método iterativo de Chebyshev es:

k	0	1	2	3
x_{1_k}	1,0	1,0224	1,0217	1,0217
x_{2_k}	1,0	1,0742	1,0731	1,0731
x_{3_k}	1,0	1,1261	1,1257	1,1257
x_{4_k}	1,0	1,1699	1,1697	1,1697
x_{5_k}	1,0	1,1992	1,2030	1,2030
x_{6_k}	1,0	1,2224	1,2264	1,2264
x_{7_k}	1,0	1,2362	1,2415	1,2415
x_{8_k}	1,0	1,2432	1,2494	1,2494
$E-T$		0,003026	$4{,}1 \times 10^{-18}$	0,0

En este caso notamos que el cambio realizado en el algoritmo si mejora el error del Teorema y el cual es una gran utilidad.

5. Conclusiones

- Se dice que el método de Newton en su versión general se denomina el método de Newton Kantorovich, permite su aplicación a ecuaciones definidas entre espacios de funciones, como son el caso de ecuaciones diferenciales o ecuaciones integrales.
- La transformación cuadrática aplicada nos da un buen augurio su utilidad cuando se desea mejorar la solución aproximada que se desea obtener para un problema dado, ya que el método iterativo de Chebyshev nos garantiza que la solución existe y es único en la región abierta y convexa que se trabaja.
- Se mensiona al Teorema de Kantorovich donde existe una contribución de las herramientas del Análisis Funcional a los problemas que se resuelven usando el Análisis Numérico que se aplica en los diversos problemas donde interactuan los sistemas de ecuaciones no lineales y cuya justificación recae en las herramientas del Análisis Funcional, así como las aplicaciones lineales, derivadas de Fréchet, etc.
- El sistema de relaciones recurrentes tiene la ventaja de reducir el problema original a una forma más simple con funciones y sucesiones escalares, y proporciona condiciones suficientes para asegurar la convergencia semilocal del método en espacios de Banach.
- Se intentó debilitar la función F al pasar de convexa a cuasi-convexa, y se concluyo que no es posible, lo cual nos indica que el Teorema de Convergencia del método iterativo de Chebyshev es fuerte.
- En la aplicación se noto que el sistema de relaciones recurrentes tiene la ventaja de que reduce el problema original a una forma más simple con funciones y sucesiones escalares, y proporciona condiciones suficientes para asegurar la convergencia semilocal del método en espacios de Banach.
- Al considerar métodos de orden superior, así como el método iterativo de Chebyshev que es de segundo orden se necesitan dos parámetros para controlar la convergencia semilocal y a partir de las relaciones recurrentes generan cuatro sucesiones reales como el orden de convergencia del método iterativo menos uno. Esto simplifica en gran manera su aplicación práctica para resolver una ecuación no lineal en espacios de Banach.
- Los métodos de orden superior no son considerados su utilidad por muchos autores, debido a su alto costo computacional, además que se tiene que evaluar la derivada segunda de Fréchet, pero nuestros resultados muestran que el costo computacional es compenzado por la velocidad de convergencia.
- En el problema que incluye a una ecuación integral, para el caso en particular se logra obtener el sistema de ecuaciones no lineal utilizando la fórmula de Gauss-Legendre, el cual es de suma importancia para lograr obtener la solución por ambos métodos.
- Es importante concluir con respecto a lo anterior, si no se conoce el sistema de ecuaciones no lineal solo se podrá analizar la convergencia o no del problema en estudio, ya que para el Teorema de Kantorovich es suficiente conocer los valores de r_1 y r_2 y en el Teorema de Chebyshev es observar que la sucesión de $\{b_n\}$ converga a 0.
- La nueva condición de parada del algoritmo dado en los métodos resulta ser una alternativa en tenerlo en consideración ya que en algunos problemas es más efectivo que en otros casos, en conclusión, es importante que los algoritmos que presentamos en el espacio de Banach muestran otra alternativa de detener la ejecución del algoritmo en ser considerado y su utilidad dependerá del problema en estudio.
- Al analizar la convergencia del Teorema de Kantorovich en el segundo problema de aplicación observamos que el error absoluto y el error del teorema difieren en su resultando, siendo más eficiente el error absoluto como se muestra en la tabla siguiente:

n	$Error$	$Error\,Teorema$
0		
1	0,24030885011662972	0,26519710951725117
2	0,00913016968850622	0,02170405257287755
3	0,00000949638598025	0,00014734725160875
4	0,00000000000818812	0,00000000679179379

- Al analizar la convergencia del Teorema de Chebyshev en el segundo problema de aplicación observamos que el error absoluto y el error del teorema difieren su resultando, siendo más eficiente el error del teorema como se muestra en la tabla siguiente:

n	$Error$	$Error\,Teorema$	$r \cdot \eta$
0			0,27009247
1	0,2432198237	0,02835991888267824	0,29845239
2	0,0062275507	0,00031962337547674	0,29877202
3	0,0000021944	0,00000004064346764	0,29877206
4	0,0000000009	0,00000000000000066	0,29877206

cabe mencionar que el valor de $r \cdot \eta$ es el radio de la bola. en el cual se encuentran los valores del vector solución.
- Se concluye la utilización de variar el η con los nuevos no es solo porque mejora el error del método sino también porque reduce el número de operaciones como se muestra en la tabla siguiente con respecto al segundo ejemplo de aplicación:

n	$Error$	$Error\,Teorema$	$r \cdot \eta$
0			0,27009247342964288
1	0,24321982	0,00302583	0,27311829953822525
2	0,00622755	0,00000004	0,27311834063765383
3	0,00000219	0,00000000	0,27311834063765383

• Es importante tener presente que una desventaja del método iterativo de Chebychev es que se tiene que conocer las dos derivadas de Fréchet para poder dar solución al problema planteado y es debido que se puede no conocer la ecuación por la forma como se presenta el problema en estudio, siendo el caso de los problemas que contiene una ecuación integral.

6. Sugerencia

Siendo concientes que en la vida diaria hay problemas de sistema de ecuaciones no lineales por resolver cada vez con mejor presición lo cual motiva a seguir analizando y investigando la importancia de los diferentes métodos que existen y que continuaran creandose con sus propias dificultades pero que son muy importantes su utilidad hoy en día.

1. Argyros I.K. Chen D, Proyecciones 12(2);119-128, 1993.
2. Argyros I.K. and Szidarovszky F,"The theory and applications of iteration methods", C.R.C. Press, Boca Raton, Fla., 1993.
3. Gutiérrez J.M, J. Comput. Appl. Math. **79** (1997), 131-145.
4. Huang Z, J. Comput. Appl. Math. **47** (1993) 211-217.
5. 'Kantorovich L.V. Akilov G.P,"Functional Analysis", Pergamon Press. Oxford, 1982.
6. Ortega J.M. Rheinboldt W.C, "Iterative Solution of Nonlinear Equations in Several Variables", Academic Press. New York, 1970.
7. Rall L.B, Computational Solution of Nonlinear Operator Equations, Krieger. Huntington. NJ, 1979.
8. Yamamoto T, Numer Math 49:203-220, 1986.
9. Candela V, "Estimadores a priori del error para los métodos iterativos de Halley y de Chebyshev", Universitat de València, España, 1988.
10. Candela V. Marquina A, Computing 44:169-184, 1990.
11. Candela V. Marquina A, Computing 45:355-367, 1990.

Revisión del método Híbrido de Alto Orden para un problema elíptico de transmisión interior

Rommel Bustinza[†] y Jonathan Munguia La Cotera[‡]
Departamento de Ingeniería Matemática & Centro de Investigación en Ingeniería Matemática (CI²MA), Universidad de Concepción, Casilla 160-C, Concepción, Chile;
Universidad Nacional de Ingeniería, Av. Túpac Amaru 210, Rímac, Lima, Perú;
[†]rbustinz@ing-mat.udec.cl, [‡]jmunguial@uni.edu.pe

Recibido el 12 de noviembre de 2020; aceptado: 22 de diciembre de 2020

En este trabajo, se aproxima la solución de un problema de transmisión de flujos entre dos sustancias a través de una interfaz interior y una condición de frontera de tipo Neumann en la frontera exterior. Utilizamos para su aproximación el ya conocido método Híbrido de Alto Orden (con sus siglas en inglés HHO), que puede clasificarse como un método de Elementos Finitos (EF) de tipo Galerkin Discontinuo (GD). Describimos el modelo matemático, su formulación variacional a nivel continuo y el esquema HHO respectivo. Discutimos existencia y unicidad y presentamos resultados del orden de convergencia óptima de la solución potencial con respecto a una norma de energía y a la norma L^2. Finalmente, se proporcionan algunos ensayos numéricos, cuyos resultados están de acuerdo con la teoría descrita.
Palabras Claves: Problema elíptico de transmisión interior, Método híbrido de alto orden.

In this work, we approximate the solution of a transmission problem of the flux between two substances through an inner interface and a Neumann boundary condition at the exterior boundary. For that approximation, we use the well- known Hybrid High Order (HHO) method, which can be classified as a Discontinuous Galerkin (DG) Finite Element (FE) method. We describe the mathematical model, its variational formulation at the continuous level, and the respective HHO scheme. We discuss the existence and uniqueness of solution and present results of the optimal convergence order of the potential with respect to an energy norm and the L^2-norm. Finally, we provided some numerical tests, which are in agreement with the theory described.

Keywords: Interior transmission elliptic problem, generalized meshes, hybrid high order method.

1. Introducción

Los problemas de transmisión (también llamados problemas de interfaz), aparecen por ejemplo en la interacción de sólidos y fluidos [1, 2]. En el caso de dispositivos electromagnéticos compuestos por diferentes materiales [3], el flujo incompresible de dos fases presenta saltos en la presión y el gradiente de presión a través de su interfaz [4].

En este artículo consideramos dos dominios disjuntos. Sea Ω_1 un dominio acotado y simplemente conexo de $\mathbb{R}^d$, $d \in 2, 3$, con frontera Lipschitz continua $\Gamma_1 := \partial\Omega_1$ y Ω_2 la región anular delimitada por Γ_1 y una segunda curva Lipschitz continua Γ_2, que está estrictamente contenida en $\mathbb{R}^d - \overline{\Omega}_1$ (ver Figura 1). Nosotros utilizaremos el método HHO que ya describimos para el problema de difusión variable [5, 6]. Para una mayor comprensión y otras aplicaciones del método HHO, sugerimos al lector revisar [7]. Nuestro problema es un problema de transmisión interior con mallas ajustadas a la interfaz. Es decir, la interfaz no corta ningún elemento de la malla. Esto nos permite manejar con comodidad interfaces poligonales e imponer con precisión las condiciones de salto sobre la interfaz. Además de extender las técnicas estándar del método HHO para problemas de difusión, pudiendo acoplar sin mucha dificultad las soluciones de cada subdominio. Aprovechar las bondades del método HHO, como por ejemplo, para manejar mallas bastante generales con elementos poligonales y nodos colgantes; polinomios de alto orden y superconvergencia.

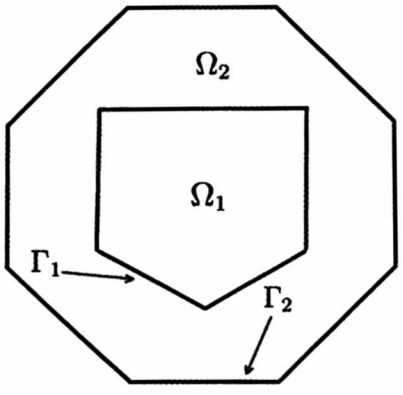


Figura 1. Geometría del problema.

En esta revisión, se realiza una ligera modificación del esquema HHO propuesto en [8]. También se ha abordado interfaces curvas que no se ajustan a los mallados de los subdominios usando el método HHO con un enfoque distinto [9], el cual deriva en un análisis a priori más complicado.

Este trabajo se organiza de la siguiente manera. En la Sección 2, se introduce el problema modelo, su formulación variacional y el estudio de su existencia y

unicidad. Los grados de libertad, los espacios discretos y los operadores discretos reconstructivos, son introducidos en la Sección 3, mientras que en la Sección 4 se obtiene la formulación discreta mixta según la filosofía HHO y se prueba que está bien puesto. El análisis de error a priori es descrito en la Sección 5, y finalmente en la Sección 6 se muestran algunos resultados numéricos.

2. Problema modelo

Consideramos el siguiente problema de transmisión: Encontrar $u_1 : \Omega_1 \to \mathbb{R}$ y $u_2 : \Omega_2 \to \mathbb{R}$ tal que

$$-\Delta u_1 = f_1 \text{ in } \Omega_1, \tag{1a}$$
$$-\Delta u_2 = f_2 \text{ in } \Omega_2, \tag{1b}$$
$$u_1 - u_2 = g \text{ on } \Gamma_1, \tag{1c}$$
$$\nabla u_1 \cdot \mathbf{n}_1 + \nabla u_2 \cdot \mathbf{n}_2 = g_1 \text{ on } \Gamma_1, \tag{1d}$$
$$\nabla u_2 \cdot \mathbf{n}_2 = g_2 \text{ on } \Gamma_2, \tag{1e}$$
$$\int_{\Omega_1} u_1 + \int_{\Omega_1} u_2 = 0, \tag{1f}$$

donde $f_1 \in L^2(\Omega_1)$ y $f_2 \in L^2(\Omega_2)$ representan los términos fuentes, $g \in H^{1/2}(\Gamma_1)$ el salto de traza del potencial sobre Γ_1, $g_1 \in H^{-1/2}(\Gamma_1)$ el salto de la componente normal del flujo sobre Γ_1 y $g_2 \in H^{-1/2}(\Gamma_2)$ la componente normal del flujo sobre Γ_2. $\mathbf{n}_1$ denota el vector normal unitario exterior a Γ_1 y $\mathbf{n}_2$ el vector normal unitario exterior a la frontera de Ω_2, dada por $\partial\Omega_2 := \Gamma_1 \cup \Gamma_2$. Por abuso de notación, denotamos la traza con su misma función. Imponemos la siguiente condición de compatibilidad, para garantizar que el problema esté bien puesto:

$$\int_{\Omega_1} f_1 + \int_{\Omega_2} f_2 + \int_{\Gamma_1} g_1 + \int_{\Gamma_2} g_2 = 0. \tag{2}$$

El punto de partida del método HHO es encontrar la formulación variacional mixta de (1). Así que, para cualquier subconjunto conexo $X \subset \mathbb{R}^d$ con medida de Lebesgue distinta de cero, denotaremos al producto interno y a la norma del espacio de Lebesgue $L^2(X)$ por $(\cdot,\cdot)_X$ y $\|\cdot\|_{0,X}$, respectivamente. También, denotamos por $Q := H^{-1/2}(\Gamma_1)$ al espacio dual de $H^{1/2}(\Gamma_1)$ con su norma dual

$$\|f^*\|_{-1/2,\Gamma_1} := \sup_{\substack{f \in H^{1/2}(\Gamma_1) \\ f \neq 0}} \frac{\langle f^*, f\rangle_{\Gamma_1}}{\|f\|_{1/2,\Gamma_1}} \quad \forall f^* \in H^{-1/2}(\Gamma_1), \tag{3}$$

donde $\langle\cdot,\cdot\rangle_{\Gamma_1}$ representa el emparejamiento dual de $H^{-1/2}(\Gamma_1)$ y $H^{1/2}(\Gamma_1)$ con respecto al producto interno $L^2(\Gamma_1)$, y análogamente para $\langle\cdot,\cdot\rangle_{\Gamma_2}$.

Ahora introducimos el espacio de Hilbert

$$\mathbf{U} := \{(v_1, v_2) \in H^1(\Omega_1) \times H^1(\Omega_2) : (v_1, 1)_{\Omega_1} + (v_2, 1)_{\Omega_2} = 0\}, \tag{4}$$

dotado de la norma $||(u_1,u_2)||^2_{\mathbf{U}} := ||u_1||^2_{1,\Omega_1} + ||u_2||^2_{1,\Omega_2}$. Al aplicar integración por partes a las ecuaciones de (1) y considerando la variable auxiliar $\xi := \nabla u_2 \cdot \mathbf{n}_2 \in H^{-1/2}(\Gamma_1)$, se obtiene la formulación variacional, que se lee: Encontrar $((u_1,u_2),\xi) \in \mathbf{U} \times Q$ tal que

$$a((u_1,u_2),(v_1,v_2)) + b((v_1,v_2),\xi) = F((v_1,v_2)), \tag{5a}$$
$$b((u_1,u_2),\lambda) = G(\lambda) \quad \forall (v_1,v_2) \in \mathbf{U},\ \forall \lambda \in Q, \tag{5b}$$

donde $a : \mathbf{U} \times \mathbf{U} \to \mathbb{R}$ y $b : \mathbf{U} \times Q \to \mathbb{R}$ son formas bilineales definidas como

$$a((u_1,u_2),(v_1,v_2)) := (\nabla u_1, \nabla v_1)_{\Omega_1} + (\nabla u_2, \nabla v_2)_{\Omega_2},$$
$$b((v_1,v_2),\xi) := \langle \xi, v_1 - v_2\rangle_{\Gamma_1}.$$

$F : \mathbf{U} \to \mathbb{R}$ y $G : Q \to \mathbb{R}$ son funcionales lineales definidas como

$$F((v_1,v_2)) := (f_1,v_1)_{\Omega_1} + (f_2,v_2)_{\Omega_2} + \langle g_1, v_1\rangle_{\Gamma_1} + \langle g_2, v_2\rangle_{\Gamma_2},$$
$$G(\lambda) := \langle \lambda, g\rangle_{\Gamma_1}.$$

Podemos definir el operador lineal acotado $\mathbf{B} : \mathbf{U} \to Q$ inducido por la forma bilineal b, al notar que

$$\begin{aligned} b((v_1,v_2),\xi) = \langle \xi, v_1 - v_2\rangle_{\Gamma_1} &= \langle \mathcal{R}(\xi), v_1 - v_2\rangle_{1/2,\Gamma_1} \\ &= \langle \mathcal{R}^*(v_1 - v_2), \xi\rangle_{-1/2,\Gamma_1}, \end{aligned}$$

donde $\mathcal{R} : H^{-1/2}(\Gamma_1) \to H^{1/2}(\Gamma_1)$ es una aplicación de Riesz, $\mathcal{R}^* : H^{1/2}(\Gamma_1) \to H^{-1/2}(\Gamma_1)$ el operador adjunto de Riesz, y $\langle\cdot,\cdot\rangle_{r,\Gamma_1}$ el producto interno sobre $H^r(\Gamma_1)$, $r \in \{-1/2, 1/2\}$. Entonces, definimos para todo $(v_1,v_2) \in \mathbf{U}$,

$$\mathbf{B}((v_1,v_2)) := \mathcal{R}^*(v_1 - v_2). \tag{6}$$

Gracias a la biyectividad de $\mathcal{R}^*$, podemos definir el núcleo de $\mathbf{B}$, como

$$V := \operatorname{Ker}\mathbf{B} = \{(v_1,v_2) \in \mathbf{U} : v_1 = v_2 \text{ on } \Gamma_1\}. \tag{7}$$

Ahora establecemos que la formulación variacional (5) tiene solución única.

Proposición 2.1 (Bien puesto). *El problema continuo (5) está bien puesto.*

Demostración. Se deduce luego de aplicar la teoría de Babuška-Brezzi. Se demuestra que las formas bilineales a y b son acotados, que a es V-elíptica y finalmente la condición inf-sup de b. El análisis es similar a la prueba del Teorema 2.1 en [8]. □

3. Marco discreto

Consideramos las notaciones utilizadas en [5, 6] para discretizar un dominio de $\mathbb{R}^d$, así como también las definiciones del operador potencial reconstructivo y de reducción local. Utilizaremos el subíndice $i = 1,2$ para referirnos a cada subdominio. Por simplicidad, consideraremos $g_1 \in L^2(\Gamma_1)$, $g_2 \in L^2(\Omega_2)$, $\Omega := \Omega_1 \cup \Gamma_1 \cup \Omega_2$, $\mathcal{T}_h$ una malla de Ω, satisfaciendo para cada $T \in \mathcal{T}_h$, que $T \subset \overline{\Omega}_1$ o $T \subset \overline{\Omega}_2$, y que no existen nodos colgantes sobre la frontera de transmisión Γ_1. A continuación, introducimos las mallas inducidas por $\mathcal{T}_h$, sobre cada subdominio $\overline{\Omega}_i$, $i \in \{1,2\}$, esto es

$$\mathcal{T}_{i,h} := \{T \in \mathcal{T}_h : T \subset \overline{\Omega}_i\},$$

Además, denotamos por $\Gamma_{1,h}$ y $\Gamma_{2,h}$, las particiones de Γ_1 y Γ_2, respectivamente, inducidas por $\mathcal{T}_{2,h}$.
Dado $k \geq 0$ el grado polinomial, definimos los grados de libertad global sobre cada subdominio:

$$\underline{\mathbf{U}}_{\mathcal{T}_{i,h}}^{k} := \left(\prod_{T \in \mathcal{T}_{i,h}} \mathbb{P}_d^k(T) \right) \times \left(\prod_{F \in \mathcal{F}_{i,h}} \mathbb{P}_{d-1}^k(F) \right). \quad (8)$$

donde $\mathcal{F}_{i,h}$ colecta las caras correspondientes a $\mathcal{T}_{i,h}$. Sea $\underline{\mathbf{v}}_{i,h} := ((v_{i,T})_{T \in \mathcal{T}_{i,h}}, (v_{i,F})_{F \in \mathcal{F}_{i,h}}) \in \underline{\mathbf{U}}_{\mathcal{T}_{i,h}}^{k}$ un elemento del subespacio discreto, Además, la restricción de $\underline{\mathbf{v}}_{i,h}$ a cada elemento $T \in \mathcal{T}_{i,h}$ es $\underline{\mathbf{v}}_{i,T} := (v_{i,T}, (v_{i,F})_{F \in \mathcal{F}_T}) \in \underline{\mathbf{U}}_{i,T}^{k} := \mathbb{P}_d^k(T) \times \left(\prod_{F \in \mathcal{F}_T} \mathbb{P}_{d-1}^k(F) \right)$, donde $i = 1, 2$. En caso no haya ambigüedad, se omitirá el subíndice i.
Ahora, establecemos nuestros espacios de aproximación discretos:

$$\underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} := \Big\{ \underline{\mathbf{v}}_h := (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}) \in \underline{\mathbf{U}}_{\mathcal{T}_{1,h}}^{k} \times \underline{\mathbf{U}}_{\mathcal{T}_{2,h}}^{k} : \sum_{i=1}^{2} \sum_{T \in \mathcal{T}_{i,h}} (v_{i,T}, 1)_T = 0 \Big\}, \quad (9)$$

$$Q_h^k := \mathbb{P}_{d-1}^k(\Gamma_{1,h}) := \prod_{F \in \Gamma_{1,h}} \mathbb{P}_{d-1}^k(F). \quad (10)$$

De aquí en adelante, adoptamos la siguiente notación: Dado $\lambda_h \in Q_h^k$, denotamos $\lambda_F := \lambda_h|_F$ para todo $F \in \Gamma_{1,h}$. Entonces, introducimos la caracterización $\lambda_h := (\lambda_F)_{F \in \Gamma_{1,h}}$. El espacio Q_h^k es provisto con la norma L^2 ponderada:

$$\|\lambda_h\|_{\Gamma_{1,h}}^2 := \sum_{F \in \Gamma_{1,h}} h_F \|\lambda_F\|_{0,F}^2, \quad \forall \lambda_h \in Q_h^k. \quad (11)$$

Introducimos también la seminorma $|||\cdot|||_h : \underline{\mathbf{U}}_{\mathcal{T}_{1,h}}^{k} \times \underline{\mathbf{U}}_{\mathcal{T}_{2,h}}^{k} \to \mathbb{R}$, la cual es dada, para cada $(\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}) \in \underline{\mathbf{U}}_{\mathcal{T}_{1,h}}^{k} \times \underline{\mathbf{U}}_{\mathcal{T}_{2,h}}^{k}$, por

$$|||(\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})|||_h^2 := \|\underline{\mathbf{v}}_{1,h}\|_{1,\mathcal{T}_{1,h}}^2 + \|\underline{\mathbf{v}}_{2,h}\|_{1,\mathcal{T}_{2,h}}^2 + \sum_{F \in \Gamma_{1,h}} h_F^{-1} \|v_{1,F} - v_{2,F}\|_{0,F}^2. \quad (12)$$

Proposición 3.1. *La aplicación $|||\cdot|||_h$ define una norma sobre $\underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}$.*

Demostración. Ver Proposición 4.1 de [8]. □

Para cada $T \in \mathcal{T}_h$, se define el operador de reducción local $\underline{\mathbf{I}}_T^k : H^1(T) \to \underline{\mathbf{U}}_T^k$ tal que, para todo $v \in H^1(T)$,

$$\underline{\mathbf{I}}_T^k v := (\pi_T^k v, (\pi_F^k v)_{F \in \mathcal{F}_T}), \quad (13)$$

donde π_T^k y π_F^k denotan las proyecciones ortogonales de L^2 sobre $\mathbb{P}_d^k(T)$ y $\mathbb{P}_{d-1}^k(F)$, respectivamente.
El correspondiente operador de reducción global $\underline{\mathbf{I}}_{\mathcal{T}_{i,h}}^k : H^1(\Omega_i) \to \underline{\mathbf{U}}_{\mathcal{T}_{i,h}}^{k}$ es definido por

$$\underline{\mathbf{I}}_{\mathcal{T}_{i,h}}^k v := ((\pi_T^k v)_{T \in \mathcal{T}_{i,h}}, (\pi_F^k v)_{F \in \mathcal{F}_{i,h}}) \quad \forall v \in H^1(\Omega_i). \quad (14)$$

Para cada $T \in \mathcal{T}_h$, se define el **operador gradiente reconstructivo local** $G_T^k : \underline{\mathbf{U}}_T^k \to \nabla \mathbb{P}_d^{k+1}(T)$ tal que, para cada $\underline{\boldsymbol{v}}_T := (v_T, (v_F)_{F \in \mathcal{F}_T}) \in \underline{\mathbf{U}}_T^k$ y cada $w \in \mathbb{P}_d^{k+1}(T)$,

$$(G_T^k \underline{\boldsymbol{v}}_T, \nabla w)_T = (\nabla v_T, \nabla w)_T + \sum_{F \in \mathcal{F}_T} (v_F - v_T, \nabla w \cdot \boldsymbol{n}_{TF})_F. \quad (15)$$

Definimos también el operador potencial reconstructivo $p_T^{k+1} : \underline{\mathbf{U}}_T^k \to \mathbb{P}_d^{k+1}(T)$ tal que, para todo $\underline{\boldsymbol{v}}_T \in \underline{\mathbf{U}}_T^k$,

$$\nabla p_T^{k+1} \underline{\boldsymbol{v}}_T := G_T^k \underline{\boldsymbol{v}}_T, \qquad \int_T p_T^{k+1} \underline{\boldsymbol{v}}_T := \int_T v_T. \quad (16)$$

Observación 3.1. *Utilizaremos propiedades de aproximación del operador $p_T^{k+1} \underline{\mathbf{I}}_T^k$ aplicado a funciones con regularidad en $H^{q+1+\delta}(T)$, para $q \in \{0, 1, \cdots, k\}$ y $\delta \in (1/2, 1]$, lo cual nos permitirá realizar el análisis a priori. Ver Apéndice A para más detalle.*

4. Formulación discreta del problema

Para formular el esquema HHO, definimos la siguiente forma bilineal $a_h : \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \times \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \to \mathbb{R}$ dada por

$$a_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) := A_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) + j_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})). \quad (17)$$

donde el término de consistencia $A_h : \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \times \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \to \mathbb{R}$ y el término de estabilidad $j_h : \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \times \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \to \mathbb{R}$ son definidos como

$$A_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) := \sum_{T \in \mathcal{T}_h} \left(G_T^k \underline{\mathbf{u}}_T, G_T^k \underline{\mathbf{v}}_T \right)_T, \quad (18)$$

$$j_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) := \sum_{T \in \mathcal{T}_h} j_T(\underline{\mathbf{u}}_T, \underline{\mathbf{v}}_T), \quad (19)$$

siendo

$$j_T(\underline{\mathbf{u}}_T, \underline{\mathbf{v}}_T) := \sum_{F \in \mathcal{F}_T} h_F^{-1} (\pi_F^k (u_F - R_T^{k+1} \underline{\mathbf{u}}_T), \pi_F^k (v_F - R_T^{k+1} \underline{\mathbf{v}}_T))_F,$$

con

$$R_T^{k+1} \underline{\mathbf{v}}_T := v_T + (p_T^{k+1} \underline{\mathbf{v}}_T - \pi_T^k p_T^{k+1} \underline{\mathbf{v}}_T).$$

También, introducimos la forma bilineal $b_h : \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \times Q_h^k \to \mathbb{R}$, la cual está definida como

$$b_h((\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}), \lambda_h) := \sum_{F \in \Gamma_{1,h}} (\lambda_F, v_{1,F} - v_{2,F})_F \quad \forall (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}) \in \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}, \ \forall \lambda_h \in Q_h^k. \quad (20)$$

Entonces el esquema discreto mixto HHO asociado a (5) se lee: Encontrar $((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), \xi_h) \in \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \times Q_h^k$ tal que

$$a_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) + b_h((\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}), \xi_h) = F_h((\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h})) \quad \forall (\underline{\mathbf{v}}_{1,h}, \underline{\mathbf{v}}_{2,h}) \in \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}, \quad (21a)$$

$$b_h((\underline{\mathbf{u}}_{1,h}, \underline{\mathbf{u}}_{2,h}), \lambda_h) = G_h(\lambda_h) \quad \forall \lambda_h \in Q_h^k, \quad (21b)$$

donde las funcionales lineales discretas F_h y G_h se definen como:

$$F_h(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h}) := \sum_{T\in\mathcal{T}_{1,h}}(f_1,v_{1,T})_T + \sum_{S\in\mathcal{T}_{2,h}}(f_2,v_{2,S})_S + \sum_{F\in\Gamma_{1,h}}(g_1,v_{1,F})_F + \sum_{F\in\Gamma_{2,h}}(g_2,v_{2,F})_F, \tag{22}$$

$$G_h(\lambda_h) := (\lambda_h,g)_{\Gamma_{1,h}} := \sum_{F\in\Gamma_{1,h}}(\lambda_F,g)_F. \tag{23}$$

Observación 4.1. *El operador lineal* $\mathbf{B}_h : \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0} \to Q_h^k$, *inducido por* b_h, *es caracterizado por*

$$\mathbf{B}_h(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h}) := (v_{1,F}-v_{2,F})_{F\in\Gamma_{1,h}} \quad \forall\,(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\in\underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}. \tag{24}$$

Observación 4.2. *La forma bilineal* a_h *induce otra seminorma sobre* $\underline{\mathbf{U}}_{\mathcal{T}_{1,h}}^k \times \underline{\mathbf{U}}_{\mathcal{T}_{2,h}}^k$, *que es dada por*

$$\|(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\|_{a,h}^2 := a_h((\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h}),(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})) \quad \forall\,(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\in\underline{\mathbf{U}}_{\mathcal{T}_{1,h}}^k\times\underline{\mathbf{U}}_{\mathcal{T}_{2,h}}^k. \tag{25}$$

Introduciendo $\mathbf{V}_h := \mathrm{Ker}(\mathbf{B}_h)$, establecemos el siguiente resultado.

Lema 4.1 (Elipticidad). a_h *is* $\mathbf{V}_h$*-elíptica.*

Demostración. Ver Lema 4.4 de [8]. □

Proposición 4.1 (Bien puesto). *El problema discreto (21) está bien puesto.*

Demostración. Se sigue de la elipticidad de a_h y la suryectividad de B_h, lo cual verifica las hipótesis de la teoría de Babuška-Brezzi. El análisis es similar a la prueba de la Proposición 4.2 en [8]. □

El siguiente resultado será util para establecer el análisis a priori.

Corolario 4.1. *Existe* $\eta > 1$, *independiente del tamaño de malla* h, *tal que*

$$\eta^{-1}|||(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})|||_h \le \|(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\|_{a,h} \quad \forall\,(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\in\mathbf{V}_h, \tag{26}$$

$$\|(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\|_{a,h} \le \eta|||(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})|||_h \quad \forall\,(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\in\underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}. \tag{27}$$

5. Análisis de Error

En esta sección, asumimos que la solución exacta $u_i \in H^{1+\delta_i}(\Omega_i)$, para algún $\delta_i \in (1/2,1]$ y $\Delta u_i \in L^2(\Omega)$ para $i\in\{1,2\}$. Estas suposiciones nos permite considerar $\xi \in L^2(\Gamma_1)$, $g_1 \in L^2(\Gamma_1)$ y $g_2 \in L^2(\Gamma_2)$. Además, denotamos la interpolación de la solución exacta por $\widehat{\underline{\mathbf{u}}}_{i,h} := \underline{\mathbf{I}}_{\mathcal{T}_{i,h}}^k u_i \in \underline{\mathbf{U}}_{\mathcal{T}_{i,h}}^k$ para $i\in\{1,2\}$. También consideramos la interpolación de la variable auxiliar ξ como $\widehat{\xi}_h \in Q_h^k$ tal que $\widehat{\xi}_h|_F := \pi_F^k(\xi)$, para cada $F\in\Gamma_{1,h}$. No es difícil deducir que $(\widehat{\underline{\mathbf{u}}}_{1,h},\widehat{\underline{\mathbf{u}}}_{2,h}) \in \underline{\mathbf{U}}_{\mathcal{T}_h}^{k,0}$. Consideramos el ensamblaje de las componentes volumétricas de la solución discreta, es decir, $v_{i,h} \in \mathbb{P}_d^k(\mathcal{T}_{i,h})$ tal que $v_{i,h}|_T = v_{i,T}\ \ \forall T\in\mathcal{T}_{i,h}$. A continuación, presentamos estimaciones de error del potencial usando la norma de energía y la norma L^2, bajo suposiciones de regularidad adicionales.

Lema 5.1 (Consistencia). *Sea* $(u_1,u_2)\in\mathbf{U}$ *la solución potencial de* (5) *y* $(\underline{\mathbf{u}}_{1,h},\underline{\mathbf{u}}_{2,h}) \in \underline{\mathbf{U}}_h^{k,0}$ *la solución potencial discreta de* (21). *Entonces, existe una constante* $C>0$, *independiente de* h, *tal que*

$$\|(\widehat{\underline{\mathbf{u}}}_{1,h},\widehat{\underline{\mathbf{u}}}_{2,h}) - (\underline{\mathbf{u}}_{1,h},\underline{\mathbf{u}}_{2,h})\|_{a,h} \le C \sup_{\substack{(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\in\underline{\mathbf{U}}_h^{k,0}\\ \|(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})\|_{a,h}=1}} \mathcal{E}_h((\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})), \tag{28}$$

donde $\mathcal{E}_h(\cdot)$ *representa el error de consistencia, y viene dado por*

$$\mathcal{E}_h((\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})) := a_h((\widehat{\underline{\mathbf{u}}}_{1,h},\widehat{\underline{\mathbf{u}}}_{2,h}),(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})) + (\mathbf{B}_h(\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h}),\widehat{\xi}_h)_{\Gamma_1} - F_h((\underline{\mathbf{v}}_{1,h},\underline{\mathbf{v}}_{2,h})) \tag{29}$$

Demostración. Se sigue directamente de las formulaciones variacionales continua (5) y discreta (21) y de (24). □

Teorema 5.1 (Estimación del error en la norma de energía). *Bajo las suposiciones del Lema 5.1, y asumiendo la regularidad adicional* $(u_1,u_2)\in H^{k+1+\delta_1}(\Omega_1)\times H^{k+1+\delta_2}(\Omega_2)$, $\exists C>0$, *independiente de* h, *tal que*

$$\|(\widehat{\underline{\mathbf{u}}}_{1,h},\widehat{\underline{\mathbf{u}}}_{2,h}) - (\underline{\mathbf{u}}_{1,h},\underline{\mathbf{u}}_{2,h})\|_{a,h} + \|\widehat{\xi}_h - \xi_h\|_{\Gamma_{1,h}} \le Ch^k\left(h_1^{2\delta_1}\|u_1\|_{k+1+\delta_1,\Omega_1}^2 + h_2^{2\delta_2}\|u_2\|_{k+1+\delta_2,\Omega_2}^2\right)^{1/2}, \tag{30}$$

donde $h := \max\{h_1,h_2\}$ *y* $h_i = \max_{T\in\mathcal{T}_{i,h}} h_T$, $i = 1,2$. *Además, aplicando propiedades de aproximación al operador* $p_T^{k+1}\underline{\mathbf{I}}_T^k$ *para cada* $T\in\mathcal{T}_h$, *se cumple también*

$$\sum_{i=1}^{2}\sum_{T\in\mathcal{T}_{i,h}} ||\nabla u_i - \nabla p_T^{k+1}\underline{\mathbf{u}}_{i,T}||_{0,T}^2 \le Ch^k\left(h_1^{2\delta_1}\|u_1\|_{k+1+\delta_1,\Omega_1}^2 + h_2^{2\delta_2}\|u_2\|_{k+1+\delta_2,\Omega_2}^2\right)^{1/2}. \tag{31}$$

Demostración. Se sigue del Lema 5.1, al acotar el error de consistencia de manera similar a las técnicas usadas en [6]. Por ejemplo, aplicando integración por parte a $f_i = -\Delta u_i$, las condiciones de transmisión y de frontera y las propiedades de aproximación del operador $p_T^{k+1}\underline{\mathbf{I}}_T^k$. □

Para la estimación del error potencial en norma L^2, consideramos el siguiente problema auxiliar: Encontrar $z\in H^1(\Omega)\cap L_0^2(\Omega)$ tal que

$$\begin{aligned} -\Delta z &= w \text{ in } \Omega := \Omega_1\cup\Gamma_1\cup\Omega_2,\\ \frac{\partial z}{\partial \boldsymbol{n}_2} &= 0 \text{ on } \Gamma_2 := \partial\Omega, \end{aligned} \tag{32}$$

con $w\in L_0^2(\Omega)$ tal que $w_i := w|_{\Omega_i}$ y $z_i := z|_{\Omega_i}$ para $i=1,2$. Luego, asumimos regularidad adicional sobre z, la solución débil de (32), de modo que $z\in H^2(\Omega)\cap L_0^2(\Omega)$

y existe $C > 0$, independiente del tamaño de malla, tal que

$$||z||^2_{2,\Omega} \leq C\,||w||^2_{0,\Omega}\,,$$

o equivalentemente

$$\|z_1\|^2_{2,\Omega_1} + \|z_2\|^2_{2,\Omega_2} \leq C\left(\|w_1\|^2_{0,\Omega_1} + \|w_2\|^2_{0,\Omega_2}\right). \quad (33)$$

Observamos que esta suposición se cumple cuando, por ejemplo, el dominio Ω es convexo. Este problema auxiliar nos permite establecer la estimación del error de la solución interpolada y la solución discreta volumétrica, con respecto a la norma L^2.

Teorema 5.2 (Estimación del error en la norma L^2). *Bajo las suposiciones del Teorema 5.1 y la condición de regularidad* (33) *del problema auxiliar. Existe $C > 0$, independiente de h, tal que para $k \geq 1$, se cumple*

$$\|\pi^k_{\mathcal{T}_{1,h}} u_1 - u_{1,h}\|_{0,\Omega_1} + \|\pi^k_{\mathcal{T}_{2,h}} u_2 - u_{2,h}\|_{0,\Omega_2} \quad (34)$$

$$\leq Ch^{k+1}\left(\sum_{i=1}^{2} h_i^{2\delta_i}\|u_i\|^2_{k+1+\delta_i,\Omega_i}\right)^{1/2}. \quad (35)$$

Para $k = 0$, asumimos que $f_i \in H^1(\mathcal{T}_{h,i})$ y $g_i \in \mathbb{P}^0_{d-1}(\Gamma_{i,h})$ para $i = 1, 2$. Entonces

$$\|\pi^0_{\mathcal{T}_{1,h}} u_1 - u_{1,h}\|_{0,\Omega_1} + \|\pi^0_{\mathcal{T}_{2,h}} u_2 - u_{2,h}\|_{0,\Omega_2} \leq$$

$$C\left[h\left(\sum_{i=1}^{2} h_i^{2\delta_i}\|u_i\|^2_{1+\delta_i,\Omega_i)}\right)^{1/2} + h^2\left(\sum_{i=1}^{2}\|f_i\|^2_{1,\mathcal{T}_{i,h}}\right)^{1/2}\right]. \quad (36)$$

Demostración. Se procede de manera similar al Teorema 5.2 de [8]. □

Observación 5.1. *Es posible encontrar el mismo orden de convergencia del Teorema* (5.2) *para la solución discreta reconstruida a través del operador potencial reconstructivo y su solución exacta, con respecto a la norma L^2.*

6. Resultados numéricos

En esta sección presentamos tres ejemplos numéricos para evaluar las propiedades de nuestro método. Utilizamos diferentes mallas para las pruebas numéricas.

Nuestro código es una extensión del desarrollado para [5, 6]. Además, calculamos el orden experimental de convergencia (r) como

$$r = \log(e_{\mathcal{T}_1}/e_{\mathcal{T}_2})/\log(h_{\mathcal{T}_1}/h_{\mathcal{T}_2})\,,$$

donde $e_{\mathcal{T}_1}$ y $e_{\mathcal{T}_2}$ son los errores asociados a las variables correspondientes considerando dos tamaño de malla consecutivos $h_{\mathcal{T}_1}$ y $h_{\mathcal{T}_2}$, respectivamente.

6.1. Ejemplo 1: Solución exacta regular

Resolvemos el problema de transmisión (1) con subdominios $\Omega_1 := (1,2)^2$ y $\Omega_2 := (0,3)^2 \setminus \overline{\Omega}_1$, usando la malla inicial de la Figura 2, tal que la solución exacta viene dada por

$$u_1(x,y) = \sin(\pi x)\sin(\pi y) - 4/\pi^2,$$
$$u_2(x,y) = \cos(\pi x)\cos(\pi y),$$

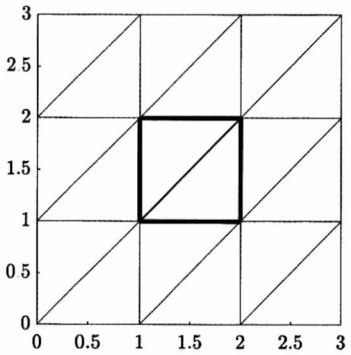


Figura 2. Malla inicial para el Ejemplo 1.

Notamos que en este caso, g_1 y g_2 no son homogéneas en Γ_1. La Figura 3 muestra los órdenes de convergencia del error al aproximar la solución potencial con polinomios a trozos de grado máximo $k \in \{0,1,2,3,4\}$. Se observa que la tasa de convergencia para este error es de $k+2$, mejor que la indicada en el Teorema 5.2. En la Figura 4 se aprecian los órdenes de convergencia del error de la aproximación del flujo cuya tasa es de $k+1$, en concordancia con el Teorema 5.1. Con respecto a la variable auxiliar de transmisión ξ, en la Figura 5 se observan los correspondientes órdenes de convergencia de la proyección de $\xi - \xi_h$ con respecto a una norma L^2 ponderada. Éstos están cercanos a $k + 3/2$ (que es $1/2$ más rápida que la predicha por el Teorema 5.1).

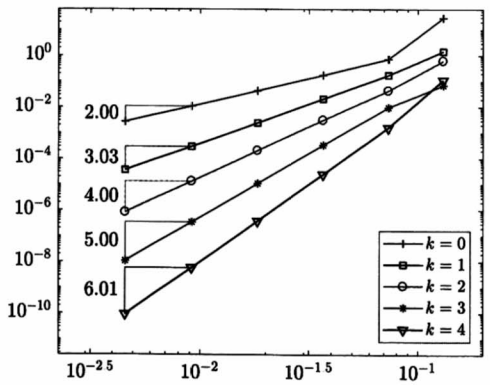


Figura 3. Error del potencial con respecto a la norma L^2 vs. h.

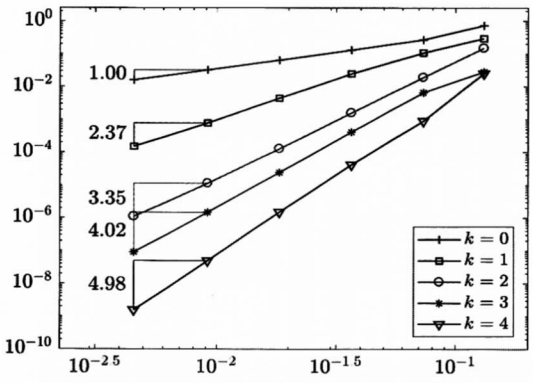


Figura 4. Error del flujo con respecto a la norma L^2 vs. h.

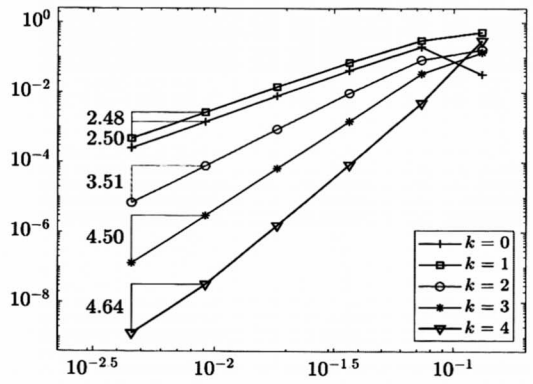


Figura 5. $||\widehat{\xi}_h - \xi_h||_{0,\Gamma_1}$ vs. h.

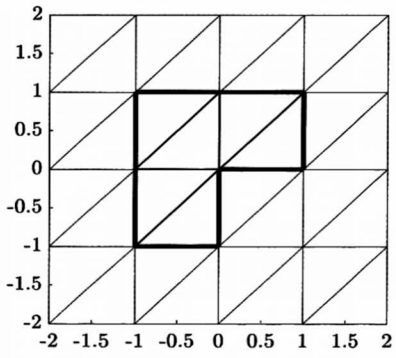


Figura 6. Malla inicial para el Ejemplo 1.

En las Figuras 7 y 8 se exhibe el comportamiento del error del potencial y del flujo con respecto al tamaño de malla h. Estos resultados no contradicen los Teoremas 5.1 y 5.2, ya que en este caso la función u_1 no es regular. Similar comportamiento se observa en la Figura 9 para el error de $\widehat{\xi}_h - \xi_h$ con respecto a la norma L^2 ponderada, con tasa de convergencia de 2/3.

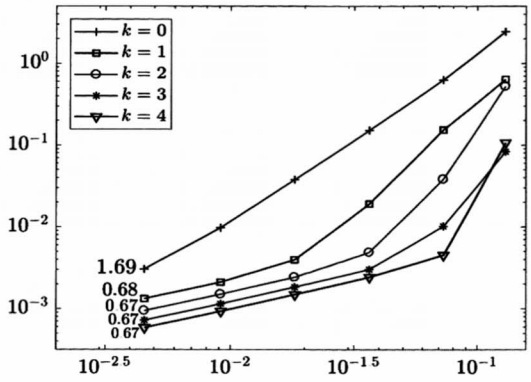


Figura 7. Error del potencial con respecto a la norma L^2 vs. h.

6.2. Ejemplo 2: Solución exacta no suave

Ahora, resolvemos un problema de transmisión, considerando $\Omega_1 = (-1,1)^2 \setminus [0,1] \times [-1,0]$ y $\Omega_2 := (-2,2)^2 \setminus \overline{\Omega}_1$ (vea Figura 6), mientras que los datos son tales que la solución exacta es (u_1, u_2), donde

$$u_1(r,\theta) = r^{2/3}\sin(2\theta/3) - c_1 \quad \text{(en coordenadas polares)},$$
$$u_2(x,y) = \sin(\pi x)\sin(\pi y) - c_2\,,$$

con c_1 y c_2 constantes reales tal que $u_j \in L^2_0(\Omega_j)$, $j \in \{1,2\}$. Señalamos que $u_1 \in H^{1+\frac{2}{3}-s}(\Omega_1)$, para un pequeño número arbitrario $s > 0$, y u_2 es una función suave.

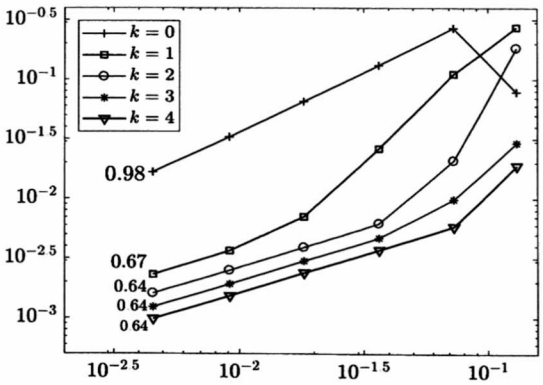


Figura 8. Error del flujo con respecto a la norma L^2 vs. h.

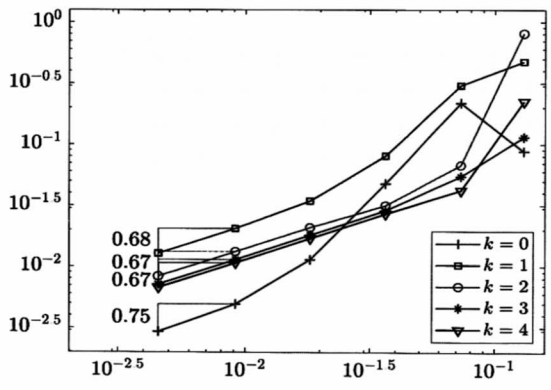


Figura 9. $||\widehat{\xi}_h - \xi_h||_{0,\Gamma_1}$ vs. h.

7. Conclusiones

En este artículo hemos propuesto una nueva formulación mixta HHO para aproximar la solución de un problema elíptico de transmisión interior, con condiciones de frontera de transmisión no homogéneas. En primer lugar, derivamos la formulación variacional, a nivel continuo, introduciendo la traza normal en la frontera de transmisión, de la solución que vive en el subdominio anular, como incógnita auxiliar. En la práctica, esta incógnita actúa como un multiplicador de Lagrange. Luego, proponemos un esquema variacional discreto, aplicando el enfoque HHO. Aunque hemos considerado, por simplicidad, una familia de mallas simpliciales uniformes para el análisis actual, es posible extenderla para tratar mallas politopales más generales.

Hemos probado que nuestro esquema mixto discreto HHO está bien puesto y presentado resultados de convergencia en la norma de energía (cf. (25)), así como en la norma usual L^2. Nuestras estimaciones de error a priori establecen que cuando aproximamos la solución con polinomios a trozos de grado máximo $k \geq 0$, al flujo con respecto a la norma L^2 y a la interpolación del potencial con respecto a la norma de energía, observamos que el error tiende a cero con un orden de convergencia $k + \delta$ (cf. Teorema 5.1). Mientras que el orden de error de la aproximación de la proyección ortogonal L^2 del potencial se comportan como $\mathcal{O}(k + 1 + \delta)$, para algún $\delta \in (1/2, 1]$ (cf. Teorema 5.2). El error de la aproximación de la proyección ortogonal L^2 del multiplicador de Lagrange se mide con respecto a una norma L^2 ponderada adecuada (cf. (11)), y converge, al menos, con el orden $k + \delta$ (cf. Teorema 5.1).

En el primer ejemplo, presentamos una solución exacta con suficiente regularidad, la cual muestra que se alcanzan las tasas óptimas de convergencia (véase Figuras 3, 4 y 5). Por otra parte, dado que la solución exacta del segundo ejemplo no es regular, no se espera obtener tasas óptimas de convergencia. Esto lo observamos en las Figuras 7, 8 y 9. Como consecuencia, esto nos motiva a desarrollar un análisis de error a posteriori, para mejorar la calidad de la aproximación y recuperar la tasa óptima de convergencia, si es posible. Esto sería tema de un trabajo futuro.
Finalmente, señalamos que el análisis descrito en este trabajo, puede ser aplicado y/o extendido para tratar problemas de transmisión lineal con difusión variable, y/o con otro tipo de condiciones de contorno en la frontera externa Γ_2. Además, teniendo en cuenta [10], estamos motivados para extender este enfoque para tratar con cierta clase de problemas de transmisión no lineal.

APÉNDICE A

En el siguiente lema, se muestra las propiedades de aproximación del operador $p_T^{k+1}\underline{\mathbf{I}}_T^k$ para funciones con regularidad en $H^{q+1+\delta}(T)$, para $q \in \{0, 1, \cdots, k\}$ y $\delta \in (1/2, 1]$.

Lema 7.1 (Propiedades de aproximación para $p_T^{k+1}\underline{\mathbf{I}}_T^k$). *Sean $k \geq 0$ el grado polinomial, $q \in \{0, \cdots, k\}$ un entero y $\delta \in (1/2, 1]$ dados. Existe un número real $C > 0$, que depende de la regularidad de la malla, depending on the mesh regularity parameter, posiblemente de d, k, q y δ, pero independiente de h_T, tal que, para todo $h \in \mathcal{H}$, para todo $T \in \mathcal{T}_h$, y para todo $v \in H^{q+1+\delta}(T)$, se cumple:*

$$\left\|v - p_T^{k+1}\underline{\mathbf{I}}_T^k v\right\|_{0,T} + h_T^{1/2}\left\|v - p_T^{k+1}\underline{\mathbf{I}}_T^k v\right\|_{0,\partial T} + h_T\left\|\nabla(v - p_T^{k+1}\underline{\mathbf{I}}_T^k v)\right\|_{0,T} + h_T^{3/2}\left\|\nabla(v - p_T^{k+1}\underline{\mathbf{I}}_T^k v)\right\|_{0,\partial T} \leq C h_T^{q+1+\delta}\|v\|_{q+1+\delta,T}. \quad (37)$$

Demostración. Ver Lema 3.3 de [8]. □

Agradecimientos

R. Bustinza ha sido parcialmente apoyado por CONICYT-Chile a través del Proyecto AFB170001 del Programa PIA: Concurso Apoyo a Centros Científicos y Tecnológicos de Excelencia con Financiamiento Basal, por el proyecto VRID-Enlace No. 218.013.044-1.0, Universidad de Concepción, y por el Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción (Chile). J. Munguia desea expresar su gratitud por el apoyo económico a CONCYTEC-Perú a través del Proyecto FONDECYT "Programas de Doctorado en Universidades Peruanas" CG-176-2015, al Instituto de Matemática y Ciencias Afines (IMCA) y a la Universidad Nacional de Ingeniería (Lima-Perú).

1. C. J. Luke and P. A. Martin., SIAM J. Appl. Math., 55(4): 904–922, 1995.
2. G. C. Hsiao and N. Nigam., Adv. Differential Equations, 8(11):1281–1318, 2003.
3. B. Heise., Impact Comput. Sci. Engrg., 5 (1993): 75–110, 1993.
4. M. Oevermann and R. Klein., J. Comput. Phys., 219(2):749–769, 2006.
5. R. Bustinza and J. Munguia. REVCIUNI 21(1): 6–14, 2018.
6. R. Bustinza and J. Munguia-La-Cotera. Numer. Methods Partial Differ. Equ., 36(3): 524–551, 2019.
7. D. A. Di Pietro and J. Droniou. Volumen 19 of Modeling, Simulation and Applications series. Springer International Publishing, 2020. 528 pages.
8. R. Bustinza and J. Munguia., Centro de Investigación en Ingeniería Matemática, Universidad de Concepción, Chile, Pre-print 2020-10, 2020.
9. E. Burman and A. Ern., SIAM J. Numer. Anal., 56(3): 1525?1546, 2017.
10. R. Bustinza and J. Munguia. *An a priori error analysis for a class of nonlinear elliptic problems with the hybrid high-order method.* Centro de Investigación en Ingeniería Matemática, Universidad de Concepción, Chile, Pre-print 2020-08, 2020.

Rompimiento de Simetría y Generación de Masa de los Bosones Escalares Exóticos en un Modelo Simétrico. Left -Right con Simetría de Gauge $SU(2)_R \otimes SU(2)_L \otimes U(1)_{B-L} \otimes \mathcal{P}$

Henry José Diaz Chávez[1†], Orlando Pereyra Ravinez[1††]
[1] *Facultad de Ciencias, Universidad Nacional de Ingeniera (UNI), Av. Tupac Amaru 210, Rimac, Lima, c.p. 15333, Perú.*
[†] *hdiaz@uni.edu.pe,*[††] *opereyra@uni.edu.pe*

Recibido el 06 de abril de 2020; aceptado: 23 de julio de 2020

La busqueda de una nueva Física, como se le llama a las diversas extensiones del Modelo Estandar (ME) de la física de partículas, nos motiva ha extender el grupo de simetría Electrodébil de $SU(2)_L \otimes U(1)_Y$ al grupo de gauge con simetría left-right $SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L} \otimes \mathcal{P}$, donde $\mathcal{P}$ representa una simetría discreta de paridad tal que las constantes de acoplamiento izquierdo-derecho satisfazen $g_L = g_R$. Este modelo representa una de las extensiones llamadas mínimas del ME, y que de acuerdo a la jerarquía en el rompimiento de la simetría (condiciones que deben cumplir los valores de expectación del vacío introducidos en el modelo) nos permita obtener el ME. El objetivo del presente trabajo es identificar al bosón de Higgs del ME, considerando un potencial escalar mas general que debe respetar todas las simetrías (gauge, discretas e invariante de Lorentz) establecidas en el modelo. Para ello se tomará en cuenta parte de lo estudiado en artículos previos, acerca de las condiciones de jerarquía y ciertas aproximaciones que deben cumplir los valores de expectación del vacío para obtener simplicidad en el desarrollo de los cálculos.

Palabras Claves: Modelos Izquierdo-Derecho, Extensiones del Modelo Estandar, simetrías, Partículas Exóticas.

The search for a new Physics, as the various extensions of the Standard Model (SM) of particle physics are called, motivates us to extend the electroweak symmetry group from $SU(2)_L \otimes U(1)_Y$ to the gauge group with left-right symmetry $SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L} \otimes \mathcal{P}$, where $\mathcal{P}$ is a parity discrete symmetry such that left-right coupling constants satisfied $g_L = g_R$. This model represents one of the minimal extensions of the SM, and which according to the hierarchy in the symmetry breaking (conditions that must be met by the vacuum expectation values introduced in the model) allowing to obtain the SM. The aim of this work is to identify the Higgs boson of the SM, considering a more general scalar potential that must respect all the symmetries (gauge, discrete symmetries and must be Lorentz invariant) established in the model. We will take into account what was studied in previous articles, about the hierarchy conditions and certain approximations that must satisfy the vacuum expectation values for obtaining greater simplicity in the development of calculations.

Keywords: Left-Right models, Standard Model Extenstions, symmetries, Exotic particles.

1 Introducción

El Modelo Estandar (ME) de las particulas fundamentales se divide en dos partes, según sea la interacción a estudiar, la Electrodébil y la Fuerte. Este modelo elaborado por Weinberg-Salam-Glashow[1] asocia a cada partícula conocida con un campo cuántico dando cuenta de sus interacciones. Todas las partículas predichas por el ME han sido observadas experimentalmente, siendo la observación final la del bosón de Higgs en 2012 por ATLAS y CMS[2][3]. A pesar del éxito del ME, la teoría no explica ciertos hechos [4]. En el ME, los neutrinos son no masivos, sin embargo, los experimentos han demostrado que los neutrinos podrían tener masas al igual como sucede con los kaones[5] presentar el fenómeno de oscilación[6]. Desde el punto de vista de los campos escalares nada nos garantiza que el bosón de Higgs predicho por el ME sea el único. Muchas de las extensiones del ME contemplan la existencia de más de una partícula escalar masiva de carga cero[7], siendo la de menor masa la del bosón de Higgs[2][3] ($M_{Higgs} = 125.10 \pm 0.14$ GeV [8]).

Este trabajo tiene por objetivo estudiar el espectro de masas del sector escalar de uno de los modelos que es una extension del ME , el llamado modelo simétrico Left-Right con simetría de gauge $SU(2)_R \otimes SU(2)_L \otimes U(1)_{B-L} \otimes \mathcal{P}$ [9][10], sin considerar el sector fuerte (Cromodinámica).

Estos modelos simétricos Left-Right son estudiados principalmente con la finalidad de dar una justificación de la pequeñes de la masa del neutrino (esto debido a la evidencia experimental de la oscilación de neutrinos[6]), así como también buscar una explicación sobre la violación de la paridad a bajas energías[11], la busqueda de candidatos a ser partícula de la materia oscura[4], etc entre otras interrogantes que como es conocido no pueden ser

explicadas por el ME.
En el presente modelo, todos los fermiones son partículas de tipo Dirac, esto incluye a los neutrinos, cuyos acoplamientos con los bosones vectoriales neutros son dados en un trabajo previo[12]. El sector escalar esta formado por, mínimo dos bi-dobletes, Φ_1, Φ_2, que dan masa, a través del rompimiento espontáneo de la simetría y el mecanismo de Higgs, a los leptones, bosones y quarks respectivamente.
Adicionalmente, el modelo presenta desde el inicio una simetría de paridad, $\mathcal{P}$, por lo tanto a través de este operador de simetría discreta conjuntamente con las simetrias de gauge es que toda la densidad lagrageana manifiesta la simetría left-right, en consecuencia las constantes de acoplamiento, g_L y g_R son iguales ($g_L = g_R$)[12], para energías mayores que la escala del ME. Al efectuar el rompimiento espontáneo de la simetria, debe obtenerse primeramente el sector electrodébil del ME $(SU(2)_L \otimes U(1)_Y$, ya que el sector fuerte es el mismo que el del ME, para llegar finalmente a la simetría natural, $U(1)_Q$, que es la simetría correspondiente al campo electromagnético y por lo tanto esta relacionado a la conservación de la carga eléctrica.
En este artículo no tomaremos en cuenta los otros sectores del modelo (Sector bosónico, Sector de Yukawa[12]) debido a que solo estamos interesados en describir con bastante detalle a las particulas escalares. Cabe mencionar que los valores de expectación del vacío (VEV), k_1, k_1', k_2, k_2', v_L, y v_R, son considerados reales, donde se debe cumplir la relación de jerarquía $v_R \gg X$, siendo X cualquiera de los VEVs diferentes de v_R.
En la sección 2, se presenta un resumen del modelo left-right, considerando la forma en que se presentan los multipletes tanto en los sectores leptónicos, escalar y bosónico, este último solo se menciona, ya que a sido discutido de manera mas detallada en artículos previos[12].
En la sección 3, se definen las llamadas ecuaciones de vínculo, las cuales toman un rol importante en el cálculo del espectro de masas de los escalares.
En la sección 4, se procede a calcular las masas de las partículas escalares propuestas en el modelo. Estos cálculos son basados en la construcción de las matrices de masas de los campos neutros así como de los simplemente cargados, obteniendose como consecuencia los bosones de Goldstone y las nuevas partículas de Higgs, tal que uno de ellos debe corresponder a la del ME.
En la sección 5, se discute a través de un análisis fenomenológico la identificación del boson de Higgs del ME, dándonos restricciones adicionales de los parámetros del modelo (generalmente dichos parámetros corresponden al potencial escalar propuesto).
En la sección 6, se colocan las conclusiones establecidas como consecuencia de los obtenido en los cálculos previos.

2 El modelo simétrico left-right

2.1 Sector leptónico

Los leptones izquierdos y derechos son dobletes en la representación fundamental de los grupos de gauge locales $SU(2)_L$ y $SU(2)_R$ respectivamente:

$$L_l \equiv \tfrac{1}{2}(1-\gamma_5)\begin{pmatrix}\psi_{\nu_l}\\ \psi_l\end{pmatrix} = \begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L ; \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L ; \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L \sim (\mathbf{2}_L, \mathbf{1}_R, -1)$$

$$R_l \equiv \tfrac{1}{2}(1+\gamma_5)\begin{pmatrix}\psi_{\nu_l}\\ \psi_l\end{pmatrix} = \begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_R ; \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_R ; \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_R \sim (\mathbf{1}_L, \mathbf{2}_R, -1)$$

Los números entre paréntesis representan la tercera componente de isospin izquierdo y derecho, $T_{3L}(T_{3R})$, y el número cuántico de hipercarga $B-L$ respectivamente[13]. Estos se relacionan a través de la fórmula de Gellmann-Nishijima [14]:

$$Q = T_{3L} + T_{3R} + \frac{B-L}{2} \tag{1}$$

En el ME el número cuántico de hipercarga era un parámetro arbitrario que no tenía significado físico pero que está relacionado con la carga eléctrica que deberán tener las partículas propuestas en el modelo después de la quiebra, ahora dicho número cuántico ya no es cualquier valor sino que esta relacionado con la diferencia entre el número bariónico y el número leptónico, que desde el punto de vista experimental resulta ser invariante [9][10].

2.2 Sector Escalar

El sector escalar consiste en dos bi-dobletes que se transforman como $(\mathbf{2}, \mathbf{2}, 0)$:

$$\Phi_1 = \begin{pmatrix}\phi_1^0 & \eta_1^+\\ \phi_1^- & \eta_1^0\end{pmatrix}, \Phi_2 = \begin{pmatrix}\phi_2^0 & \eta_2^+\\ \phi_2^- & \eta_2^0\end{pmatrix}, \tag{2}$$

la carga eléctrica que se muestra en los campos escalares son justificadas en [15]. El bi-doblete Φ_1 le da masa a los leptones conocidos (incluido los neutrinos) y el otro bi-doblete, Φ_2 le da masa a los quarks. Los dobletes $\chi_L \sim (\mathbf{2}, \mathbf{1}, +1)$ y $\chi_R \sim (\mathbf{1}, \mathbf{2}, +1)$ se introducen no sólo para completar la jerarquia en el rompimiento de simetría hacia el grupo natural $U(1)_Q$ sino también para que se proteja la simetría left-right del modelo. Esta simetría de paridad queda rota explícitamente cuando el campo neutro del doblete χ_R gana un valor de expectación del vacío diferente de cero (rompimiento espontáneo de la simetría), es decir, $v_R \neq 0$.

$$\chi_L = \begin{pmatrix}\chi_L^+\\ \chi_L^0\end{pmatrix}, \quad \chi_R = \begin{pmatrix}\chi_R^+\\ \chi_R^0\end{pmatrix}, \tag{3}$$

El sector de Higgs es diferente para cada modelo que se estudia, uno establece la forma de los multipletes escalares (dependiendo lo que se desea investigar) y adicionalmente se propone un potencial escalar, lo más general, es decir, aquel que respeta las simetrías de gauge del

modelo y alguna simetría discreta adicional. Por ejemplo, para el caso de nuestro modelo, se propone el siguiente potencial[12]:

$$V = V^{(2)} + V^{(4a)} + V^{(4b)} + V^{(4c)} + V^{(4d)} + V^{(4e)} \quad (4)$$

donde:

$$\begin{aligned}
V^{(2)} &= \frac{1}{2}\sum_{i=1,2}^{2}\left[\mu_{ii}^2\, Tr(\Phi_i^\dagger\Phi_i) + H.c.\right] \\
&+ \mu_{LR}^2\left(\chi_L^\dagger\chi_L + \chi_R^\dagger\chi_R\right) \\
V^{(4a)} &= \frac{1}{2}\sum_{i=1,2}^{2}\left[\lambda_{ii} Tr(\Phi_i^\dagger\Phi_i)^2 + H.c.\right], \\
V^{(4b)} &= \frac{1}{2}\sum_{i=1,2}^{2}\lambda'_{ii}\left(Tr\Phi_i^\dagger\Phi_i\right)^2, \qquad (5)\\
V^{(4c)} &= \rho_{12}\, Tr\left(\Phi_1^\dagger\Phi_1\Phi_2^\dagger\Phi_2\right), \\
V^{(4d)} &= \frac{1}{2}\left[\sum_{i=1,2}^{2}(\Lambda_{ii}\, Tr\Phi_i^\dagger\Phi_i(\chi_L^\dagger\chi_L\chi_R^\dagger\chi_R)\right. \\
&+ \overline{\Lambda}_{ii}(\chi_L^\dagger\Phi_i\Phi_i^\dagger\chi_L + \chi_R^\dagger\Phi_i\Phi_i^\dagger\chi_R) + \\
&+ \left.\overline{\Lambda}'_{ii}(\chi_L^\dagger\tilde{\Phi}_i\tilde{\Phi}_i^\dagger\chi_L + \chi_R^\dagger\tilde{\Phi}_i\tilde{\Phi}_i^\dagger\chi_R))\right], \\
V^{(4e)} &= \lambda_{LR}\left[(\chi_L^\dagger\chi_L)^2 + (\chi_R^\dagger\chi_R)^2\right].
\end{aligned}$$

Se observan varios parámetros introducidos en el potencial que están relacionados a través de las llamadas ecuaciones de vínculo.
Como nuestro estudio se basa en éste sector escalar, es necesario suponer la existencia de los VEVs a través de los campos escalares neutros.

$$<\Phi_1>=\frac{1}{\sqrt{2}}\begin{pmatrix} k_1 & 0 \\ 0 & k'_1\end{pmatrix}, \quad <\Phi_2>=\frac{1}{\sqrt{2}}\begin{pmatrix} k_2 & 0 \\ 0 & k'_2\end{pmatrix}, \quad (6)$$

y

$$\langle\chi_L\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_L\end{pmatrix}, \quad \langle\chi_R\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_R\end{pmatrix}, \quad (7)$$

Estos VEVs son reales, ya que estamos interesados en el espectro de masa de los escalares. Cabe destacar que si estuvieramos interesados en otros estudios tal como Violación CP[11], estos VEVs los consideraríamos complejos.

2.3 Sector Bosónico

Como toda extensión del ME, el modelo propone la existencia de siete bosones vectoriales: Un bosón neutro no masivo, el fotón A_μ, dos bosones neutros masivos (Z_L, Z_R) y cuatro cargados masivos ($W_L^\pm$, $W_R^\pm$). El estudio de este sector no forma parte del objetivo de este artículo, ver [12].

3 Ecuaciones de Vínculo

La ecuaciones de vínculo se obtienen cuando el potencial lo expandimos alrededor de los valores esperados del vacío mediante el llamado rompimiento espontáneo de la simetría, es decir con la condición:

$$\frac{\partial V}{\partial\phi_i}|_{\phi=\langle\phi\rangle} = 0. \quad (8)$$

donde $\langle\phi\rangle$ es cualquier valor de expectación: k_1, k_2, k'_1, k'_2, v_L y v_R.
Tomando en cuenta la condición para el mínimo del potencial (8), se obtienen seis ecuaciones de vínculo:

$$\begin{aligned}
(a)\quad & k_1\mu_{11}^2 + (\lambda_{11}+\lambda'_{11})k_1^3 + \lambda'_{11}k_1k_1'^2 \\
&+ \frac{k_1}{2}(v_L^2+v_R^2)(\Lambda_{11}+\overline{\Lambda}'_{11}) + \frac{1}{2}k_1k_2^2\rho_{12} = 0, \\
(b)\quad & k'_1\mu_{11}^2 + (\lambda_{11}+\lambda'_{11})k_1'^3 + \lambda'_{11}k'_1k_1^2 \\
&+ \frac{k'_1}{2}(v_L^2+v_R^2)(\Lambda_{11}+\overline{\Lambda}_{11}) + \frac{1}{2}k'_1k_2'^2\rho_{12} = 0, \\
(c)\quad & k_2\mu_{22}^2 + (\lambda_{22}+\lambda'_{22})k_2^3 + \lambda'_{22}k_2k_2'^2 \\
&+ \frac{k_2}{2}(v_L^2+v_R^2)(\Lambda_{22}+\overline{\Lambda}'_{22}) + \frac{1}{2}k_2k_1^2\rho_{12} = 0, \\
(d)\quad & k'_2\mu_{22}^2 + (\lambda_{22}+\lambda'_{22})k_2'^3 + \lambda'_{22}k'_2k_2^2 \qquad (9)\\
&+ \frac{k'_2}{2}(v_L^2+v_R^2)(\Lambda_{22}+\overline{\Lambda}_{22}) + \frac{1}{2}k'_2k_1'^2\rho_{12} = 0, \\
(e)\quad & \mu_{LR}^2\, v_L + \lambda_{LR}\, v_L^3 + \frac{v_L}{2}\left(k_1'^2(\Lambda_{11}+\overline{\Lambda}_{11}) + k_2'^2\times\right. \\
& \left.(\Lambda_{22}+\overline{\Lambda}_{22}) + k_1^2(\Lambda_{11}+\overline{\Lambda}'_{11}) + k_2^2(\Lambda_{22}+\overline{\Lambda}'_{22})\right) \\
& = 0, \\
(f)\quad & \mu_{LR}^2 v_R + \lambda_{LR}v_R^3 + \frac{v_R}{2}\left(k_1'^2(\Lambda_{11}+\overline{\Lambda}_{11}) + k_2'^2\right. \\
& \left.(\Lambda_{22}+\overline{\Lambda}_{22}) + k_1^2(\Lambda_{11}+\overline{\Lambda}'_{11}) + k_2^2(\Lambda_{22}+\overline{\Lambda}'_{22})\right) = 0.
\end{aligned}$$

Estas ecuaciones representan condiciones que deben cumplir tanto los VEVs como los parámetros que fueron definidos previamente a través del potencial y se utilizarán con el fin de simplificar los elementos de matrices tanto para los campos escalares simplemente cargados y los campos neutros.
Se puede observar de estas ecuaciones que existe la opción de que cualquiera de los valores de expectación tomen, en algún momento, el valor de cero. Eso dependerá exclusivamente de lo que se esté interesado por estudiar, por ejemplo, para nuestro caso, por razones de simplicidad en la obtención de los valores de masa de estos escalares podemos aproximarlos al valor de cero comparados con el valor de expectación del vacío v_R, ver [12].

4 Matrices de Masas de los Bosones Escalares Exóticos

Expandiendo los campos escalares neutros alrededor de los valores esperados mínimos (VEVs reales) del potencial, y considerándolos (a dichos campos) complejos, los dos bi-dobletes y dobletes escalares toman la forma:

Para los bi-dobletes:

$$\Phi_1 = \begin{pmatrix} \frac{1}{\sqrt{2}}(k_1 + H_{1a} + i\,I_{1a}) & \eta_1^+ \\ \phi_1^- & \frac{1}{\sqrt{2}}(k_1' + H_{2a} + i\,I_{2a}) \end{pmatrix}, \tag{10}$$

$$\Phi_2 = \begin{pmatrix} \frac{1}{\sqrt{2}}(k_2 + H_{1b} + i\,I_{1b}) & \eta_2^+ \\ \phi_2^- & \frac{1}{\sqrt{2}}(k_2' + H_{2b} + i\,I_{2b}) \end{pmatrix},$$

Para los dobletes:

$$\chi_L = \begin{pmatrix} \chi_L^+ \\ \frac{1}{\sqrt{2}}(v_L + H_{1L} + i\,I_{1L}) \end{pmatrix}, \tag{11}$$

$$\chi_R = \begin{pmatrix} \chi_R^+ \\ \frac{1}{\sqrt{2}}(v_R + H_{1R} + i\,I_{1R}) \end{pmatrix},$$

Observar que sólo los campos neutros son expandidos alrededor de un valor esperado mínimo (VEV) mas no los campos cargados.
En nuestro modelo no existen campos escalares doblemente cargados, ello debido a que no hemos introducido en el sector Higgs los tripletes escalares, esto implicaría un estudio de neutrinos masivos a través del mecanismo de Seesaw[16][17].

4.1 Campos Escalares Simplemente Cargados

La matriz de masa de cualquier campo escalar se obtiene a partir de los términos cuadráticos de los campos dentro del potencial escalar del modelo. Es decir, para un potencial arbitrario, la expresión general de los elementos de masa de los campos escalares se obtienen de[18] :

$$\sum_{i,j} \frac{\partial^2 V}{\partial\phi_k \partial\phi_i}|_{\phi=\langle\phi\rangle} (T^a)_{ij} \langle\phi_j\rangle = 0 \tag{12}$$

recordar que T^a son los generadores del grupo de simetría. La matriz de masa de los campos escalares (masas al cuadrado)esta dada por:

$$\mathcal{M}_{ij}^2 = \frac{\partial^2 V}{\partial\phi_k \partial\phi_i}|_{\phi=\langle\phi\rangle} \tag{13}$$

La base que consideramos en la expresión de la matriz para los campos escalares simplemente cargados es: η_1^+, ϕ_1^-, η_2^+, ϕ_2^-, χ_L^+ y χ_R^+.

$$\mathcal{M}_{SC}^2 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{16} \\ a_{21} & a_{22} & \cdots & a_{26} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61} & a_{62} & \cdots & a_{66} \end{pmatrix} \tag{14}$$

donde:

$$\begin{aligned}
a_{11} &= \frac{\rho_{12}(k_2'^2 - k_2^2)}{2} - \frac{1}{2}(v_L^2 + v_R^2)(\Lambda_{11} + \overline{\Lambda}_{11}') \\
&+ k_1'^2\lambda_{11} \\
a_{12} &= a_{21} = 0, \\
a_{13} &= a_{31} = \frac{\rho_{12}k_1k_2}{2}, \\
a_{14} &= a_{41} = a_{15} = a_{51} = 0, \\
a_{16} &= a_{61} = 0, \\
a_{22} &= \frac{\rho_{12}(k_2^2 - k_2'^2)}{2} - \frac{1}{2}(v_L^2 + v_R^2)(\Lambda_{11} + \overline{\Lambda}_{11}) \\
&+ k_1^2\lambda_{11}, \\
a_{23} &= a_{32} = 0, \\
a_{24} &= a_{42} = \frac{\rho_{12}k_1'k_2'}{2}, \\
a_{25} &= a_{26} = a_{52} = a_{62} = 0,
\end{aligned} \tag{15}$$

$$\begin{aligned}
a_{33} &= \frac{\rho_{12}(k_1'^2 - k_1^2)}{2} - \frac{1}{2}(v_L^2 + v_R^2)(\Lambda_{22} + \overline{\Lambda}_{22}') \\
&+ k_2'^2\lambda_{22} \\
a_{34} &= a_{43} = a_{35} = a_{53} = 0, \\
a_{36} &= a_{63} = 0, \\
a_{44} &= \frac{\rho_{12}(k_1^2 - k_1'^2)}{2} - \frac{1}{2}(v_L^2 + v_R^2)(\Lambda_{22} + \overline{\Lambda}_{22}) \\
&+ k_2^2\lambda_{22} \\
a_{45} &= a_{46} = a_{54} = a_{64} = 0,
\end{aligned}$$

$$\begin{aligned}
a_{55} &= -\frac{1}{2}k_1'^2(\Lambda_{11} + \overline{\Lambda}_{11}) - \frac{1}{2}k_2'^2(\Lambda_{22} + \overline{\Lambda}_{22}) \\
&- \frac{1}{2}k_1^2(\Lambda_{11} + \overline{\Lambda}_{11}') - \frac{1}{2}k_2^2(\Lambda_{22} + \overline{\Lambda}_{22}'), \\
a_{56} &= a_{65} = 0, \\
a_{66} &= -\frac{1}{2}k_1'^2(\Lambda_{11} + \overline{\Lambda}_{11}) - \frac{1}{2}k_2'^2(\Lambda_{22} + \overline{\Lambda}_{22}) \\
&- \frac{1}{2}k_1^2(\Lambda_{11} + \overline{\Lambda}_{11}') - \frac{1}{2}k_2^2(\Lambda_{22} + \overline{\Lambda}_{22}'),
\end{aligned}$$

Para construir esta matriz se han utilizado las ecuaciones de vínculo dadas en la expresión (9) con la finalidad de simplificar estos elementos de matriz. Además, se observa que es una matriz simetrica verificandose $a_{55} = a_{66}$.

El siguiente paso es diagonalizar la matriz, pero para ello se usará una aproximación para simplificar aun más los cálculos, ya que son muy extensos. Se usará la siguiente aproximación para los VEVs que se justifican en la referencia [12]:

$$k_1 \approx k_1' \approx 0, \;\; k_2 \approx k_2', \;\; v_L = 0. \tag{16}$$

Esta relaciones de los VEVs han sido consideradas en un artículo previo[12], donde la jerarquía que deben cumplir los VEVs es la siguiente: $v_R \gg k_1, k_2, k_1', k_2', v_L$. Estas aproximaciones (16), son justificadas por las ecuaciones de vínculo, donde también el hecho de hacer $v_L = 0$ no afecta a ningún sector dentro del lagrangiano[12]. Es decir, el doblete χ_L no interactúa con las partículas (fermiones) conocidas del ME, siendo, según los modelos actuales, candidato a ser partícula de la materia oscura. Por lo tanto, con las aproximaciones dadas en (16) se tiene la nueva matriz:

$$\mathcal{M}_{Cf}^2 = \begin{pmatrix} e_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & e_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & e_{66} \end{pmatrix}, \tag{17}$$

donde los elementos de la matriz diagonal son:

$$\begin{aligned} e_{11} &= -\frac{v_R^2}{2}(\Lambda_{11}+\overline{\Lambda}'_{11}), \\ e_{22} &= -\frac{v_R^2}{2}(\Lambda_{11}+\overline{\Lambda}_{11}), \\ e_{33} &= k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}'_{22}), \\ e_{44} &= k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}_{22}), \\ e_{55} &= e_{66}=-k_2^2\,(2\Lambda_{22}+\overline{\Lambda}_{22}+\overline{\Lambda}'_{22}) \end{aligned} \tag{18}$$

Por ser una matriz diagonal los valores propios resultan ser los mismos elementos de la diagonal, por lo tanto, los valores de masa de estos campos (masas al cuadrado) estan dados por:

$$\begin{aligned} m^2_{H_1^+} &= -\frac{v_R^2}{2}(\Lambda_{11}+\overline{\Lambda}'_{11}), \\ m^2_{H_2^-} &= -\frac{v_R^2}{2}(\Lambda_{11}+\overline{\Lambda}_{11}), \\ m^2_{H_3^+} &= k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}'_{22}), \\ m^2_{H_4^-} &= k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}_{22}), \\ m^2_{H_5^+} &= m^2_{H_6^+}=-k_2^2\,(2\Lambda_{22}+\overline{\Lambda}_{22}+\overline{\Lambda}'_{22}) \end{aligned} \tag{19}$$

Se observa de estos resultados que los campos escalares simplemente cargados resultan ser partículas masivas. Es conocido de la literatura que el rompimiento espontáneo de la simetría de un grupo continuo de mayor a uno de menor dimensión se efectua a partir de los campos escalares neutros donde como consecuencia aparecen los llamados Bosones de Goldstone, y no a través de los campos escalares cargados, con la finalidad de que siempre se cumpla la conservación de la carga eléctrica [19].

De las ecuaciones dadas en (19) y del hecho que según la jerarquía de los VEVs se debe cumplir: $v_R \gg X$, donde X es cualquier otro valor de expectación, se obtienen:

$$\begin{aligned} \Lambda_{11}+\overline{\Lambda}'_{11} &< 0, \quad & \Lambda_{11}+\overline{\Lambda}_{11}<0, \\ \Lambda_{22}+\overline{\Lambda}'_{22} &< 0, \quad & \Lambda_{22}+\overline{\Lambda}_{22}<0, \\ 2\Lambda_{22}+\overline{\Lambda}_{22}+\overline{\Lambda}'_{22} &< 0, & \end{aligned} \tag{20}$$

siendo estas desigualdades restricciones adicionales que deben cumplir los parámetros. También se puede mencionar que hay dos partículas simplemente cargadas que presentan degeneración, es decir, tienen la misma masa (estrictamente hablando, su cuadrado), $m^2_{H_5^+}$, $m^2_{H_6^+}$. Finalmente, obtenemos sus masas, a partir de (19), de estos escalares cargados son:

$$\begin{aligned} m_{H_1^+} &= v_R\sqrt{-\frac{1}{2}(\Lambda_{11}+\overline{\Lambda}'_{11})}, \\ m_{H_2^-} &= v_R\sqrt{-\frac{1}{2}(\Lambda_{11}+\overline{\Lambda}_{11})}, \\ m_{H_3^+} &= \sqrt{k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}'_{22})}, \\ m_{H_4^-} &= \sqrt{k_2^2\,\lambda_{22}-\frac{v_R^2}{2}(\Lambda_{22}+\overline{\Lambda}_{22})}, \\ m_{H_5^+} &= m_{H_6^+}=k_2\sqrt{-\,(2\Lambda_{22}+\overline{\Lambda}_{22}+\overline{\Lambda}'_{22})} \end{aligned} \tag{21}$$

Aproximando las masas de los escalares $m_{H_3^+}$ y $m_{H_4^-}$ para valores de v_R grandes comparado a los otros VEVs:

$$\begin{aligned} m_{H_3^+} &\approx v_R\sqrt{\frac{-(\Lambda_{22}+\overline{\Lambda}'_{22})}{2}}-\frac{k_2^2\,\lambda_{22}}{v_R\sqrt{-2(\Lambda_{22}+\overline{\Lambda}'_{22})}} \\ &+ \mathcal{O}(1/v_R^3). \\ m_{H_4^+} &\approx v_R\sqrt{\frac{-(\Lambda_{22}+\overline{\Lambda}_{22})}{2}}-\frac{k_2^2\,\lambda_{22}}{v_R\sqrt{-2(\Lambda_{22}+\overline{\Lambda}_{22})}} \\ &+ \mathcal{O}(1/v_R^3). \end{aligned}$$

Se puede deducir a partir de estos resultados que:

$$m_{H_1^+},\ m_{H_2^+},\ m_{H_3^+},\ m_{H_4^+} > m_{H_5^+}, \tag{22}$$

4.2 Campos Escalares Neutros (Parte Real)

Para hallar las masas de estos campos se consideró los términos cuadráticos en el potencial escalar referentes a: H_{1a}, H_{1b}, H_{2a}, H_{2b}, H_{L1} y H_{R1}, que a la vez representan la base con la cual se construyó la matriz de estos campos escalares neutros (operadores de campos hermitianos).

$$\mathcal{M}_h^2=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{16} \\ b_{21} & b_{22} & \cdots & b_{26} \\ \vdots & \vdots & \ddots & \vdots \\ b_{61} & b_{62} & \cdots & b_{66} \end{pmatrix} \tag{23}$$

donde los elementos de matriz son de la forma:

$$\begin{aligned} b_{11} &= k_1^2\,(\lambda_{11}+\lambda'_{11}), \\ b_{12} &= b_{21}=\frac{1}{2}\rho_{12}k_1k_2, \\ b_{13} &= b_{31}=\lambda'_{11}k_1k'_1, \\ b_{14} &= b_{41}=0, \\ b_{15} &= b_{51}=\frac{k_1v_L(\overline{\Lambda}_{11}+\overline{\Lambda}'_{11})}{2} \end{aligned}$$

$$
\begin{aligned}
b_{16} &= b_{61} = \frac{k_1 v_R(\overline{\Lambda}_{11} + \overline{\Lambda}'_{11})}{2}, \\
b_{22} &= k_2^2\,(\lambda_{22} + \lambda'_{22})\,, \\
b_{23} &= b_{32} = 0, \\
b_{24} &= b_{42} = \lambda'_{22} k'_1 k'_2, \\
b_{25} &= b_{52} = \frac{k_2 v_L(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
b_{26} &= b_{62} = \frac{k_2 v_R(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
\\
b_{33} &= k_1'^2\,(\lambda_{11} + \lambda'_{11})\,, \\
b_{34} &= b_{43} = \frac{1}{2}\rho_{12} k'_1 k'_2, \\
b_{35} &= b_{53} = \frac{k'_1 v_L(\overline{\Lambda}_{11} + \overline{\Lambda}'_{11})}{2}, \\
b_{36} &= b_{63} = \frac{k'_1 v_R(\overline{\Lambda}_{11} + \overline{\Lambda}'_{11})}{2}, \\
b_{44} &= k_2'^2\,(\lambda_{22} + \lambda'_{22})\,, \\
b_{45} &= b_{54} = \frac{k'_2 v_L(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
b_{46} &= b_{64} = \frac{k'_2 v_R(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
b_{55} &= \lambda_{LR}\, v_L^2, \\
b_{56} &= b_{65} = 0, \\
b_{66} &= \lambda_{LR}\, v_R^2,
\end{aligned}
\tag{24}
$$

Similar al caso anterior de campos escalares simplemente cargados, se ha utilizado las ecuaciones de vínculo para expresar de forma mas sencilla los elementos de matriz. Cabe mencionar la gran importancia de las ecuaciones de vínculo, no solo para obtener restricciones de los parámetros del modelo, sino también de simplificar haciendo los cálculos menos laboriosos.
Se ha considerado continuar con las aproximaciones dadas en la ecuación (16), con el fin de simplificar los cálculos. Por tanto, la nueva matriz tiene la expresión:

$$
\mathcal{M}_H^2 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & c_{22} & 0 & c_{24} & 0 & c_{26} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & c_{42} & 0 & c_{44} & 0 & c_{46} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & c_{62} & 0 & c_{64} & 0 & c_{66}
\end{pmatrix}
\tag{25}
$$

donde:

$$
\begin{aligned}
c_{22} &= b_{22} = k_2^2\,(\lambda_{22} + \lambda'_{22})\,, \\
c_{24} &= c_{42} = k_2^2 \lambda'_{22}, \\
c_{26} &= c_{62} = b_{26} = b_{62} = \frac{k_2 v_R(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
c_{44} &= k_2'^2\,(\lambda_{22} + \lambda'_{22})\,, \\
c_{46} &= c_{64} = \frac{k_2 v_R(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22})}{2}, \\
c_{66} &= b_{66} = \lambda_{LR}\, v_R^2,
\end{aligned}
\tag{26}
$$

la matriz (25) es simétrica y real. Se puede encontrar una matriz ortogonal que diagonalice a dicha matriz, sin embargo, es posible obtener los valores propios exactos sin la necesidad de proponer una matriz ortogonal. Por lo tanto, al diagonalizarla se obtuvo los siguientes valores de masa teniendo en cuenta las aproximaciones dadas en (16)(valores cuadráticos):

1. Campos no masivos:

$$
m^2_{H_1} = m^2_{H_2} = m^2_{H_3} = 0, \tag{27}
$$

se observa la existencia de tres campos de Higgs neutros sin masas.

2. Campos masivos:

$$
\begin{aligned}
m^2_{H_4} &= \frac{1}{2}\left[k_2^2(\lambda_{22} + 2\lambda'_{22}) + \lambda_{LR} v_R^2 - \sqrt{\Delta}\right], \\
m^2_{H_5} &= k_2^2\,\lambda_{22}, \\
m^2_{H_6} &= \frac{1}{2}\left[k_2^2(\lambda_{22} + 2\lambda'_{22}) + \lambda_{LR} v_R^2 + \sqrt{\Delta}\right],
\end{aligned}
\tag{28}
$$

donde:

$$
\begin{aligned}
\Delta &= \left[\lambda_{LR} v_R^2 - k_2^2(\lambda_{22} + 2\lambda'_{22})\right]^2 \\
&+ 2k_2^2\, v_R^2 \left(\overline{\Lambda}_{22} + \overline{\Lambda}'_{22}\right)^2,
\end{aligned}
\tag{29}
$$

Se observa la existencia de tres Higgs neutros masivos donde se puede decir que uno de dichos campos correspondería al bosón de Higgs del ME. Además, de las expresiones de masas, (28) se puede afirmar que el parámetro λ_{22} debe ser positivo, como fue observado en la sección anterior y que se cumple lo siguiente:

$$
m^2_{H_6} \gg m^2_{H_4}, m^2_{H_5},
$$

para valores de $v_R \gg X$, donde X representa cualquier otro VEV diferente a v_R.

4.3 Campos Escalares Neutros (Parte Imaginaria)

Tomando en cuenta las componentes imaginarios de los campos de Higgs, la representación matricial de las contribucions de los términos cuadráticos es expresada de la forma siguiente:

$$
\mathcal{M}^2_{SI} = \begin{pmatrix}
n_{11} & 0 & \cdots & 0 \\
0 & n_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & n_{66}
\end{pmatrix},
\tag{30}
$$

donde, sin tomar en cuenta las ecuaciones de vínculo, resulta ser matriz diagonal, además, sus elementos de matriz (en este caso resultados exactos) de la diagonal toman la forma:

$$
\begin{aligned}
n_{11} &= \frac{1}{4}(v_R^2 + v_L^2)(\overline{\Lambda}'_{11} + \Lambda_{11}) + \frac{k_1'^2 \lambda'_{11}}{2} \\
&+ \frac{k_2^2 \rho_{12}}{4} + \frac{k_1^2}{2}(\lambda'_{11} + \lambda_{11}) + \frac{\mu^2_{11}}{2}
\end{aligned}
$$

$$
\begin{aligned}
n_{22} &= \frac{1}{4}(v_R^2 + v_L^2)(\overline{\Lambda}'_{22} + \Lambda_{22}) + \frac{k_2'^2 \lambda'_{22}}{2} \\
&+ \frac{k_1^2 \rho_{12}}{4} + \frac{k_2^2}{2}(\lambda'_{22} + \lambda_{22}) + \frac{\mu^2_{22}}{2}
\end{aligned}
$$

$$
\begin{aligned}
n_{33} &= \frac{1}{4}(v_R^2+v_L^2)(\overline{\Lambda}_{11}+\Lambda_{11})+\frac{k_1^2\lambda'_{11}}{2} \\
&+ \frac{k_2'^2\rho_{12}}{4}+\frac{k_1'^2}{2}(\lambda'_{11}+\lambda_{11})+\frac{\mu_{11}^2}{2} \\
n_{44} &= \frac{1}{4}(v_R^2+v_L^2)(\overline{\Lambda}_{22}+\Lambda_{22})+\frac{k_2^2\lambda'_{22}}{2} \\
&+ \frac{k_1'^2\rho_{12}}{4}+\frac{k_2'^2}{2}(\lambda'_{22}+\lambda_{22})+\frac{\mu_{22}^2}{2} \\
n_{55} &= \frac{k_1'^2}{4}(\overline{\Lambda}_{11}+\Lambda_{11})+\frac{k_2'^2}{4}(\overline{\Lambda}_{22}+\Lambda_{22}) \\
&+ \frac{k_1^2}{4}(\overline{\Lambda}'_{11}+\Lambda_{11})+\frac{k_2^2}{4}(\overline{\Lambda}'_{22}+\Lambda_{22}) \\
&+ \frac{\mu_{LR}^2}{2}+\frac{\lambda_{LR}}{2}v_L^2, \qquad (31) \\
n_{66} &= \frac{k_1'^2}{4}(\overline{\Lambda}_{11}+\Lambda_{11})+\frac{k_2'^2}{4}(\overline{\Lambda}_{22}+\Lambda_{22}) \\
&+ \frac{k_1^2}{4}(\overline{\Lambda}'_{11}+\Lambda_{11})+\frac{k_2^2}{4}(\overline{\Lambda}'_{22}+\Lambda_{22}) \\
&+ \frac{\mu_{LR}^2}{2}+\frac{\lambda_{LR}}{2}v_R^2,
\end{aligned}
$$

Esta matriz es expresada en la base: $\{I_{1a}, I_{1b}, I_{2a}, I_{2b}, I_{L1}, I_{R1}\}$; por ser ésta diagonal, automáticamente obtenemos los valores de masa (los cuadrados) de estos campos:

$$
\begin{aligned}
m^2_{H_{I_1}} &= n_{11}, \quad m^2_{H_{I_2}} = n_{22}, \quad m^2_{H_{I_3}} = n_{33}, \\
m^2_{H_{I_4}} &= n_{44}, \quad m^2_{H_{I_5}} = n_{55}, \quad m^2_{H_{I_6}} = n_{66}(32)
\end{aligned}
$$

Sin embargo, al tomar en cuenta las ecuaciones de vínculo se obtiene que dichos campos no adquieren masa, es decir:

$$
m^2_{H_{I_1}} = m^2_{H_{I_2}} = m^2_{H_{I_3}} = m^2_{H_{I_4}} = m^2_{H_{I_5}} = m^2_{H_{I_6}} = 0. \qquad (33)
$$

Aquí no es necesario usar las aproximaciones de la ecuación (16). Se puede decir que se tiene 6 bosones de Goldstone.

5 Identificando al Bosón de Higgs del ME:

Tomando en cuenta los cuadrados de las masas obtenidas para los campos de Higgs reales (neutros), ecuación (28), al hallar las masas considerando valores grandes de v_R comparados a otros valores de expectación introducidos en el modelo se obtiene:

$$
\begin{aligned}
m_{H_4} &\approx k_2\sqrt{\lambda_{22}+\lambda'_{22}+\lambda'^2_{22}-\frac{(\overline{\Lambda}'_{22}+\Lambda_{22})^2}{2\lambda_{LR}}} \\
&+ \mathcal{O}(1/v_R^2) \\
m_{H_5} &= k_2\sqrt{\lambda_{22}}, \qquad (34) \\
m_{H_6} &\approx v_R\sqrt{\lambda_{LR}}+\mathcal{O}(1/v_R),
\end{aligned}
$$

el único valor exacto es el bosón escalar m_{H_5}, no necesariamente éste es el bosón de Higgs del ME. Debido a la gran cantidad de parámetros necesarios para describir el modelo, en especial en el sector de escalar y el sector de Yukawa[12], existe la posibilidad de que m_{H_4} sea el bosón de Higgs del ME (= 125 GeV)[3].
Al considerar el caso en que $v_R \to \infty$, pero finito, las masas de m_{H_4} y m_{H_6} se aproximan a:

$$
\begin{aligned}
m_{H_4} &\approx k_2\sqrt{\lambda_{22}+\lambda'_{22}+\lambda'^2_{22}-\frac{(\overline{\Lambda}'_{22}+\Lambda_{22})^2}{2\lambda_{LR}}} \\
m_{H_6} &\approx v_R\sqrt{\lambda_{LR}}, \qquad (35)
\end{aligned}
$$

se observa que el parámetro λ_{LR} tiene un signo definido (positivo), donde se cumple:

$$
m_{H_6} \gg m_{H_4},\ m_{H_5},
$$

como fue observado anteriormente.
Se puede considerar el siguiente análisis:

- Si $m_{H_6} > m_{H_4} > m_{H_5}$, entonces m_{H_5} sería el bosón de Higgs del ME.
  De acuerdo al particle data group (PDG)[8]: $M_{Higgs} = 125.10 \pm 0.14$ GeV. Por tanto, comparando:

$$
M_{Higgs} = 125.10 = M_{H_5} = k_2\sqrt{\lambda_{22}}, \qquad (36)
$$

  Se mencionó anteriormente que a escalas del ME, es decir, energías alrededor de la masa del bosón neutro Z^0, el parámetro k_2 puede tomar un valor máximo de 246 GeV [8]). De aquí se obtiene la siguiente condición para λ_{22}:

$$
\lambda_{22} \geq 0.259 \qquad (37)
$$

- Si $m_{H_6} > m_{H_5} > m_{H_4}$
  En este caso m_{H_4} sería el bosón de Higgs del ME, por lo que al comparar se tendría la siguiente condición:

$$
k_2\sqrt{\lambda_{22}+\lambda'_{22}+\lambda'^2_{22}-\frac{(\overline{\Lambda}'_{22}+\Lambda_{22})^2}{2\lambda_{LR}}} = 125.10 \qquad (38)
$$

  Como ya se mencionó a escalas de energía del ME, el máximo valor que puede tomar k_2 es 246 GeV, obteniéndose:

$$
\lambda_{22}+\lambda'_{22}+\lambda'^2_{22}-\frac{(\overline{\Lambda}'_{22}+\Lambda_{22})^2}{2\lambda_{LR}} \geq 0.067. \qquad (39)
$$

Los resultados obtenidos en las expresiones (37) y (39) se complementa con lo obtenido a través de las ecuaciones de vínculo y son útiles en los cálculos fenomenológicos. Recordar que de acuerdo a las bibliografías, ver [20], primero se rompe la simetría a través del VEV v_R, cuyo valor es del orden de 1 TeV (en general $v_R > 1$ TeV).

6 Conclusiones

El objetivo de este trabajo ha sido identificar el bosón de Higgs del ME así como calcular las masas de los demás escalares y esto lo hemos obtenido considerando ciertas condiciones adicionales que deben cumplir los parámetros correspondientes. De acuerdo a los resultados obtenidos el modelo presenta (sólo en el sector escalar): 3 bosones de Higgs neutros masivos y tres neutros no masivos, 6 partículas masivas simplemente cargadas y 6 partículas no físicas sin masas (componentes imaginarias de los campos neutros), todo ello como consecuencia del rompimiento espontáneo de la simetría. Se puede decir que de lo obtenido anteriormente las partículas no físicas representan seis bosones de Goldstone que van a ser absorbidas para dar masa a los bosones vectoriales que propone el modelo: $W_L^{\pm}$, $W_R^{\pm}$, Z_L, Z_R, (Se cumple el Teorema de Goldstone) El caso del bosón vectorial que lleva la fuerza electromagnética, el fotón, sigue siendo no masivo como debe ser para que el modelo sea consistente. Es necesario mencionar que la mayoría de los modelos se diferencian por su sector escalar, ya que es relevante definir dicho sector con el fin de aplicar el mecanismo de Higgs junto con el rompimiento espontáneo de la simetría para hacer de la teoría coherente (renormalizable), ademas, aunque no hemos enfatizado en el sector leptónico todas los fermiones predichos en el modelo son partículas de Dirac, incluyendo a los neutrinos (partículas masivas en el modelo).

El bosón de Higgs más masivo es el m_{H_6}, como consecuencia de su dependencia directamente proporcional a v_R, que es el VEV mas grande (en TeV), comparado a los otros, que a su vez son soluciones de las ecuaciones de vínculo. Por ejemplo, si $\lambda_{LR} = 16$ y $v_R = 1$ TeV entonces: $m_{H_6} = 4$ TeV.

Agradecimientos

Se agradece el apoyo de este trabajo de investigación al Programa de Becas de Doctorado en física de la Facultad de Ciencias de la UNI-Convenio Nro. 168 FONDECYT-UNI. Asimismo, se agradece el aporte invaluable del profesor Vicente Pleitez del Instituto de Fisica Teórica (IFT)- Sao Paulo - Brasil.

1. S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264
2. S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716**, 30 (2012) [arXiv:1207.7235 [hep-ex]].
3. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716**, 1 (2012) [arXiv:1207.7214 [hep-ex]].
4. E. C. F. S. Fortes, V. Pleitez and F. W. Stecker, JCAP **1802**, no. 02, 026 (2018); [aeXiv:1703.05275].
5. M. Gell-Mann, A. Pais, Phys. Rev. **97**, 1387 (1955).
6. Langacker, Paul (2010). TAKAAKI KAJITA,REVISTA BOLIVIANA DE FÍSICA 28s, 1–3, 2016 ISSN 1562–3823.DISCOVERY OF NEUTRINO OSCILLATIONS
7. J. C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974) Erratum: [Phys. Rev. D **11**, 703 (1975)]; R. N. Mohapatra and J. C. Pati, Left-Right Gauge Symmetry and an Isoconjugate Model of CP Violation, Phys. Rev. D **11**, 566 (1975); R. N. Mohapatra and J. C. Pati, A Natural Left-Right Symmetry, Phys. Rev. D **11**, 2558 (1975); G. Senjanovic and R. N. Mohapatra, Exact Left-Right Symmetry and Spontaneous Violation of Parity, Phys. Rev. D **12**, 1502 (1975); G. Senjanovic, Spontaneous Breakdown of Parity in a Class of Gauge Theories, Nucl. Phys. B **153**, 334 (1979), and references therein.
8. M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018) and 2019 update.
9. A. Maiezza, M. Nemevsek, F. Nesti and G. Senjanovic, Phys. Rev. D **82**, 055022 (2010), [arXiv:1005.5160 [hep-ph]].
10. G. Senjanovic and V. Tello, Parity and the origin of neutrino mass, arXiv:1912.13060 [hep-ph].
11. R. N. Mohapatra and G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980)
12. H. Diaz Chavez, V. Pleitez and O. P. Ravinez, Dirac neutrinos in a $SU(2)$ left-right symmetric model, arXiv:1908.02828 [hep-ph].
13. R. E. Marshak and R. N. Mohapatra, Phys. Lett. **91B**, 222 (1980).
14. A. Davidson, Phys. Rev. D **20**, 776 (1979).
15. E. Castillo-Ruiz, 1, V. Pleitez, 2 , and O. P. Ravinez 1, 3, Electric charge assignment in quantum field theories, arXiv:2001.03104v1 [hep-ph] 9 Jan 2020, [arXiv:1505.01934 [hep-ph]].
16. G. Senjanovic and V. Tello, Phys. Rev. Lett. **119**, no. 20, 201803 (2017) [arXiv:1612.05503 [hep-ph]].
17. G. Senjanovic and V. Tello, Disentangling Seesaw in the Minimal Left-Right Symmetric Model, arXiv:1812.03790 [hep-ph].
18. Rabindra N. Mohapatra, Palash B. Pal, Massive Neutrinos in Physics and Astrophysics, World Scientific Lecture Notes in Physics-Vol. 72,
19. Francis Halzen and Alan D. Martin, Quarks amd Leptons: An Introductory Course in Modern Particle Physics Jhon Wiley and Sons - 1984. [
20. Chang Hun Lee, Doctor of Philosophy, 2017, Left-Right Symmetric Model and its TeV-Scale Phenomenology, Dissertation directed by: Rabindra N. Mohapatra.

Estudio de circuitos protectores de baterías de iones de litio en el proceso de carga y descarga

César Martín Cruz Salazar[1],Ronald Nicolas Saenz Chuqui
Facultad de Ciencias, Universidad Nacional de Ingeniera (UNI), Av. Tupac Amaru 210, Rimac, Lima, c.p. 15333, Perú.
[1] *ccruz@uni.edu.pe*

Las baterías de iones de litio se utilizan ampliamente como fuente de alimentación que suministra el accionamiento eléctrico de múltiples dispositivos. Son sensibles a la operación fuera de su área de operación recomendada, lo que podría conducir a un menor tiempo de vida, daños y riesgo de explosión. Se aplica un sistema de gestión de batería (BMS) para controlar y proteger la batería de las condiciones anormales. Este trabajo de investigación consiste en medir el voltaje de corte máximo en la carga y el voltaje de corte mínimo en la descarga de baterías de iones de litio usando circuitos de protección de baterías. Se mide para el caso de una celda y para un conjunto de tres celdas en serie. En el primer caso se utilizó como protector el TP4056 y se desarrolló un circuito basado en una placa de Arduino Nano para registrar datos de voltaje de la carga y descarga cada cierto tiempo. En el segundo caso se desarrolló un circuito 3S 4P que consiste en tres celdas en serie y conectando en paralelo 3 series más de 3 celdas en serie conectados al gestor de batería BMS 3S 20Amperios.

Palabras Claves: Arduino Nano, Modulo TP-4056, BMS, carga de batería li-on, descarga de batería li-on, batería de iones de litio, li-ion, 18650.

Lithium-ion batteries are widely used as a power source that supplies the electrical drive for multiple devices. power supply for multiple devices. They are sensitive to operation outside their recommended operating range, which could lead to shorter lifetime, damage and risk of explosion. A battery management system is applied. management system (BMS) is applied to control and protect the battery from abnormal conditions. This research work is to measure the maximum cut-off voltage at charging and the minimum cut-off voltage at discharging of lithium-ion batteries using ion batteries using battery protection circuits. It is measured for the case of one cell and for a set of three cells in series. three cells in series. In the first case, the TP4056 was used as the protector and a circuit based on an Arduino Nano board was developed to register the battery. based circuit was developed to record charging and discharging voltage data every few seconds. In the second In the second case, a 3S 4P circuit was developed consisting of three cells in series and connecting in parallel 3 series of 3 more 3 cells in series connected to the BMS 3S 20Amp battery manager.

Keywords: Arduino Nano, TP4056 module, BMS, li-ion battery charge, li-ion battery discharge, lithium ion batteries, li-ion, 18650.

1. Introducción

En la actualidad, nos hemos acostumbrado a la movilidad como el uso del teléfono celular, el uso de un ordenador portátil y el uso de un medio de transporte cada vez más común, como son los patinetes eléctricos, bicicletas eléctricas, sillas de ruedas y autos eléctricos, etc.
Todos ellos dependen de un elemento que proporciona la energía necesaria para el funcionamiento de todos estos dispositivos indicados líneas arriba.
Esto es la batería de iones de litio conformada por celdas de iones de litio (li-ion para lo que sigue) (Figura. 1)

Figura 1. *Celdas 18650 de iones de litio de 3.7v.*

Para hacer frente a la exigencia común de lograr la sostenibilidad energética, disminuir los riesgos ambientales como la emisión de gases de efecto invernadero, el calentamiento global, etc., y el agotamiento de los combustibles fósiles.

Es de gran importancia reemplazar el vehículo convencional por un vehículo eléctrico que use una batería ecológica de cero emisiones eléctrica.

Con características tales como la tasa de auto descarga baja, la densidad alta de almacenamiento de energía, y el ciclo de vida útil prolongada, hacen que las baterías li-ion se hayan convertido en la principal fuente de energía de los Vehículos Eléctricos (EVs) [1]-[2].

Las celdas 18650 son uno de los tipos de baterías más comerciales y de mayor uso en el mercado. Las pilas ordinarias traen consecuencias negativas para el medio ambiente (las sustancias químicas que hay en el interior de estas son, por ejemplo, manganeso, zinc o hidróxido potásico, todos estos elementos son tóxicos, difíciles y caros de reciclar), por lo que es cada vez más habitual

disponer de baterías recargables en base a celdas li-ion.

El objetivo de este trabajo de investigación es verificar experimentalmente los voltajes de corte hacia arriba como hacia abajo de una celda li-ion usando la placa TP4056 (Fig.2) y de un conjunto de celdas li-ion usando la placa BMS 3S 20A (Fig.3).

En el segundo caso para la medición experimental sea construido una batería de 12v conectando un conjunto de celdas 18650 li-ion.

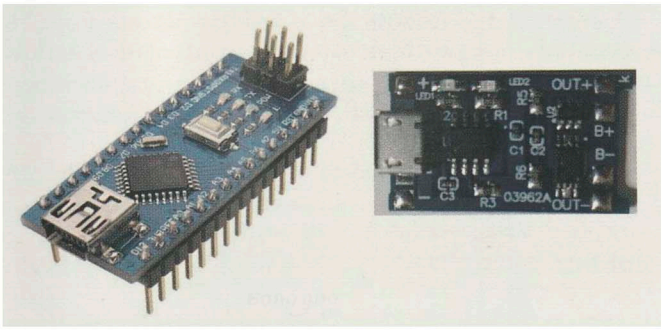


Figura 2. *Placa Arduino Nano a la izquierda y la placa de protección TP4056 a la derecha.*

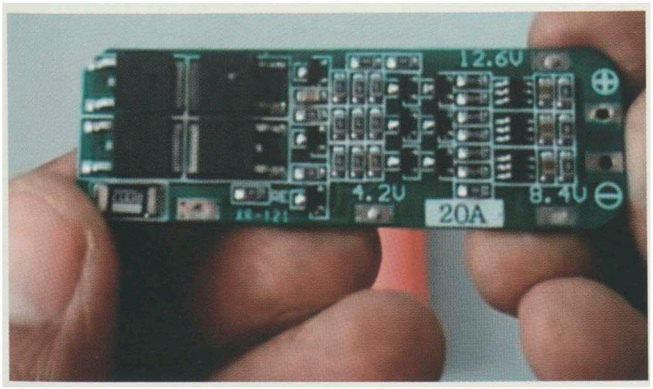


Figura 3. *Placa BMS 3S 20Amp.*

En las secciones 2 y 3, se tiene una breve explicación de la celda li-ion así como la descripción del sistema usado para la medición. En las secciones 4, 5, 6 y 7 se muestran resultados obtenidos, el BMS, resultados del circuito BMS y observaciones finales.

2. Batería li-ion

La celda li-ion es una batería donde la energía eléctrica puede ser almacenada como energía química y entonces esta energía química se convierte en energía eléctrica para cuando se requiera [3].

En la tabla 1 describimos el interior de una batería de iones de litio.

Tabla 1. Estructura de la batería li-ion.

ELEMENTO	DESCRIPCION
Catodo	Terminal positivo, esta compuesto por LiCoO2. Sustituido por el de fosfato de hierro-litio(LiFePO4).
Anodo	Terminal negativo, esta compuesto por grafito.
Separador	Barrera que evita cortocircuito entre los dos electrodos.
Electrolito	Medio situado entre cátodo y ánodo que permite el paso de la carga electrica entre ellos. Se trata de una sal de litio disuelta en un disolvente orgánico [4].

Una celda de iones de litio es una batería de alta energía en la que Li + se incrusta y escapa de materiales positivos y negativos cuando se carga y descarga.

Como se ilustra en la figura 4, de izquierda a derecha, una batería consta de un colector de corriente catódica, materiales activos de electrodo negativo, electrolito, un separador, materiales activos de electrodo positivo y un colector de corriente de ánodo. Los materiales de electrodos positivos de las baterías de iones de litio son compuestos de iones de litio, comúnmente LiCoO2, LiNiO2, LiMn2O4, LiFePO4 y LiNixCo1-2xMnxO2, y así sucesivamente.

Los materiales de electrodos negativos son comúnmente LixC6, TiS2, V2O5, etc. El electrolito es un disolvente orgánico en el que las sales de litio, como LiPF6, LiBF4, LiClO4, LiAsF6, etc., son solubles. Los disolventes son principalmente carbonato de etileno (EC), carbonato de propileno (PC), carbonato de dimetilo (DMC), carbonato de metilo de cloro (ClMC), etc.

El papel principal del separador en una celda es aislar los electrodos positivo y negativo, al tiempo que permite el transporte de iones. Recientemente, una membrana microporosa de polietileno (PE) o polipropileno (PP) se ha utilizado comercialmente como separador. Los iones de litio se desunen del compuesto del cátodo y se intercalan en la red del ánodo durante el proceso de carga. El cátodo tiene un alto potencial y un pobre estado de litio, mientras que el ánodo tiene un bajo potencial y un rico estado de litio.

Cuando se descarga, el Li + escapa del ánodo y se incrusta en el cátodo, produciendo un rico estado de litio en el cátodo. Por lo tanto, el proceso de carga y descarga de las baterías también es un proceso de desintercalación e intercalación de litio entre los dos electrodos, de ahí el nombre de "baterías de mecedoras".

Para mantener el equilibrio de carga, durante el proceso de carga y descarga, el mismo número de electrones se mueve con el Li + entre el cátodo y el ánodo a través del circuito externo. Por lo tanto, se produce una reacción redox entre el cátodo y el ánodo [9].

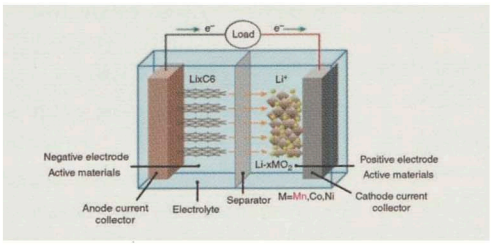


Figura 4. *Proceso de descarga de una batería Li-ion [9]*

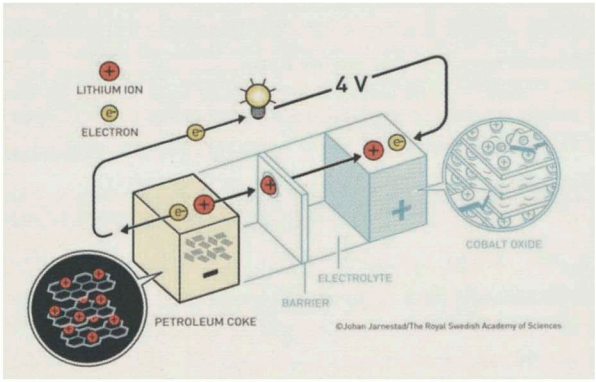


Figura 5. *Batería Li-ion Yoshino. Fuente: nobelprize.org*

La celda li-ion tiene un recubrimiento exterior de metal, que ahora es particularmente importante porque la pila está presurizada.
Esta caja de metal tiene un orificio de ventilación sensible a la presión. Si la pila se calienta tanto que corre el riesgo de explotar debido a una sobrepresión, esta ventilación se encarga de liberar la presión generada en el interior de la pila.
El respiradero está colocado como medida de seguridad. También dispone de un dispositivo que evita sobrecalentamientos [5].
La carga completa ocurre cuando la batería alcanza el umbral de voltaje y la corriente cae al 3 % de la corriente nominal. Un aumento de la corriente de carga no adelanta una carga completa. Aunque la batería alcanza el pico de voltaje más rápido, en consecuencia, la carga de saturación tardará más. Con una corriente más alta, la carga al inicio es más corta, pero la saturación después de un cierto tiempo tomará más tiempo. Una carga de corriente elevada, sin embargo, llenará rápidamente la batería a aproximadamente un 70 %. Una batería li-ion no necesita estar completamente cargada como es el caso de baterías plomo ácido, ni es deseable hacerlo. De hecho, es mejor no cargarla completamente porque un alto voltaje sobrecarga la batería. La elección de un umbral de voltaje más bajo o la eliminación de la carga de saturación por completo prolonga la vida de la batería, pero esto reduce el tiempo de ejecución. Los cargadores de otros productos van a la capacidad máxima y no se pueden ajustar; con lo que la vida útil del producto en si se percibe con menor importancia [6].

3. Materiales y metodología

3.1. Materiales

El circuito elaborado para obtener la gráfica de voltaje de la carga y descarga de la batería de iones de litio fue construido a partir de un microcontrolador Arduino Nano, una placa TP4056, un protoboard, una placa booster step-up, un circuito optoacoplador, resistencias, leds y cables de conexión.

Tabla2 .Equipo usado para el circuito con TP4056.

Nro	Equipamiento usado	Especificación
1	Arduino Nano	Entrada de 5V
2	TP4056	Entrada de 5V,1000mA
3	Booster STEP-UP	Salida de 3V-32V
4	Resistencias	7 Ω, 1000Ω, 1kΩ, 10kΩ
5	Multímetro	Mide Voltaje, Ohms, etc

EL TP4056 es un módulo de protección y además un controlador lineal completo de corriente y voltaje para una sola celda li-ion. El voltaje de carga es fijado a 4.2V. El TP4056 finaliza automáticamente el ciclo de carga cuando la corriente de carga cae a 1/10 del valor programado después de alcanzar el voltaje final. Otras características incluyen monitor de corriente, bloqueo por bajo voltaje, recarga automática y dos pines de estado para indicar la terminación de la carga y la presencia de un voltaje de entrada [7]. El circuito utilizado para las mediciones se puede visualizar en la figura 6 y figura 7. Ambos circuitos son uno solo al unirse los números que tienen puntos suspensivos. En la figura 8 se visualiza la foto del circuito armado que corresponde a las figuras 6 y 7.

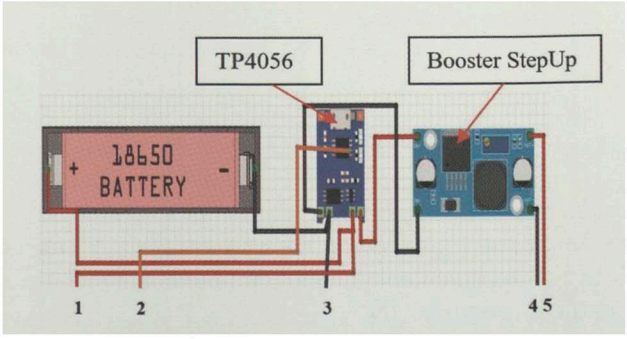


Figura 6. *Primera parte del circuito en el cual se puede visualizar el módulo TP4056 el cual controla la carga y descarga de la celda li-ion.*

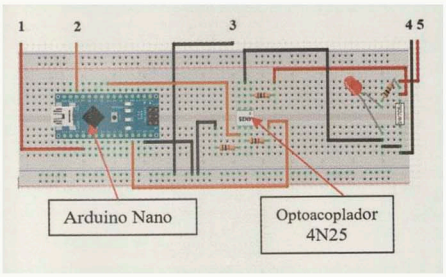


Figura 7. *Segunda parte del circuito en el cual se puede visualizar el Arduino Nano que registra voltaje en la carga y descarga.*

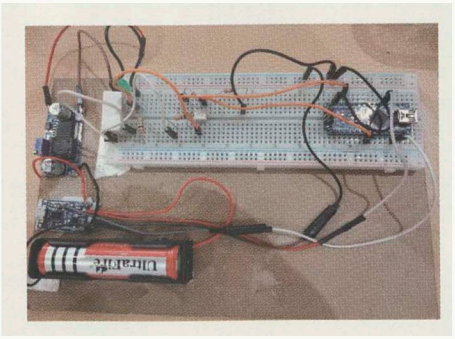


Figura 8. *Foto del circuito usado en las mediciones. Aquí no se visualiza el cable que se conecta vía USB a la computadora PC. Pero si se utiliza.*

3.2. Metodología

En la figura 6, la placa TP4056 se encarga de controlar la carga de la batería li-ion y a la vez se encarga de su protección, es decir, que no supere su carga máxima y su descarga mínima. Con cables que se conectan a los terminales IN+ e IN- de la placa TP4056 o en su defecto con un cable que se conecta al conector micro-usb se carga la celda li-ion mediante un adaptador de corriente de 5v y 1 o 2 amperios. Los cables que salen de los terminales B+ y a B- del TP4056 van al Arduino Nano (ver figura 7) a una de sus entradas analógicas, mediante lo cual se mide el voltaje, y mediante un programa hecho en Arduino podemos obtener los datos de voltaje de carga.
El voltaje de corte máximo de carga de la batería es controlado por la placa TP4056. El programa requiere un estado lógico 0, para ello se utiliza un cable soldado a su PIN6 del TP4056 y el otro extremo se conecta a una entrada digital del Arduino. El estado 0 lógico que se recibe indica que la celda li-ion está completamente cargada y es en este momento que ya no se sigue suministrando corriente a la celda li-ion.
En la figura 6, los cables de conexión conectados al OUT+ y OUT- de la placa TP4056 se conectan al IN+ e IN- del módulo booster step-up (figura 9) el cual da una salida calibrada de 5V, este voltaje se obtiene a través del OUT+ y OUT- de la placa blooster el cual se usa para descargar la celda li-ion mediante una resistencia de 7.5Ohms y 5W. De esta manera se puede obtener la gráfica de la descarga de la batería li-ion.

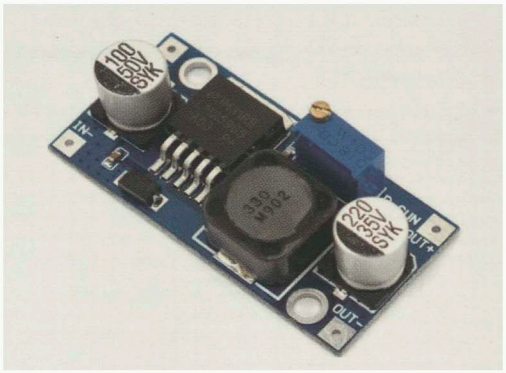


Figura 9. *Placa Booster Step Up*

Cuando se llega a la carga mínima de la batería la placa TP4056 ya no permite que siga descargándose la celda li-ion, y mediante un cable de conexión conectado en el PIN5 del optoacoplador que también se conecta en la entrada digital D5 del Arduino nano da un 0 lógico que indica que la batería está descargada. Cabe resaltar que se usó un optoacoplador para separar las tierras del Arduino nano de la placa booster step-up.

4. Análisis de gráficos

Con los datos obtenidos se obtuvo las siguientes gráficas en el proceso de carga de diferentes marcas de celdas li-ion.

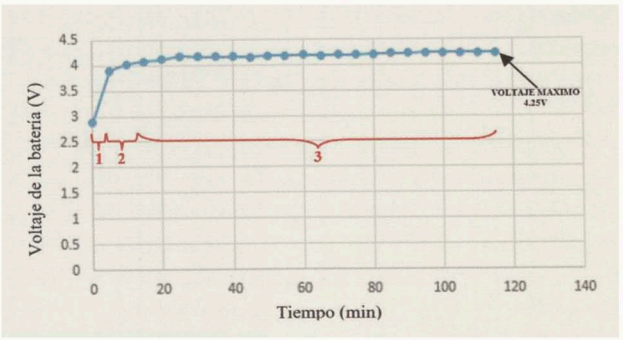


Figura 10. *Grafica de la carga de la celda li-ion de la marca CAFINI.*

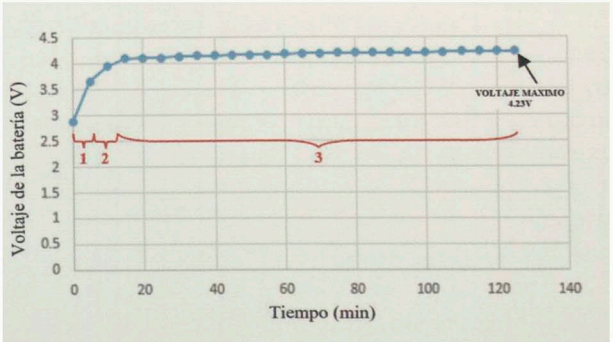


Figura 11. *Grafica de la carga de la celda li-ion de la marca ULTRAFIRE.*

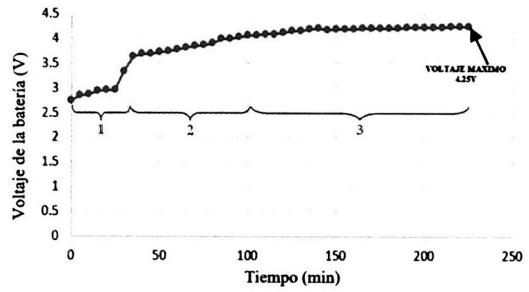


Figura 12. *Grafica de la carga de la celda li-ion de la marca LG.*

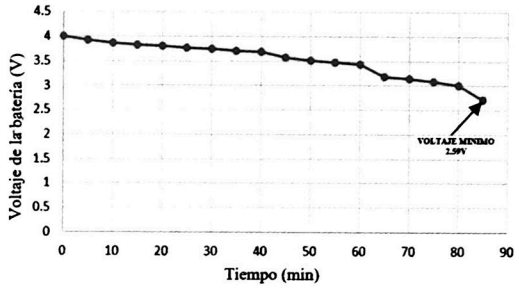


Figura 15. *Grafica de la descarga de la celda li-ion de la marca LG.*

En la figuras mostradas se puede ver que la batería presenta 3 etapas al momento de cargarla, en la primera etapa presenta una carga rápida en un corto tiempo donde el voltaje va desde su voltaje mínimo hacia el intervalo 3.5V-3.6V, en su segunda etapa empieza una carga menos rápida que la anterior donde el voltaje va desde el intervalo 3.5V-3.6V hacia el intervalo 4.0V-4.10V y en la última etapa presenta una carga mucho más lenta , el cual tiene un tiempo de carga más lento que las dos etapas anteriores hasta llegar a su voltaje máximo, se puede ver que el tiempo de carga de la primera etapa y segunda etapa juntos es igual al tiempo de la tercera etapa. Así también, se obtuvo la gráfica de descarga para cada tipo de batería usando una resistencia de 5W de 7Ohms, tomado del B+ y B- de la placa TP4056.

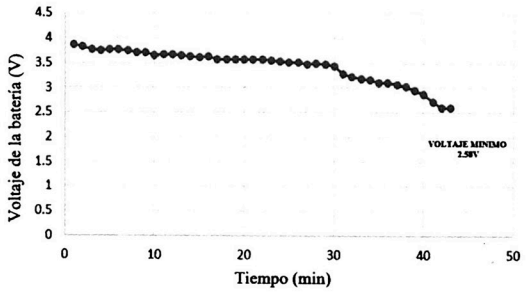


Figura 13. *Grafica de la descarga de la celda li-ion de la marca CAFINI.*

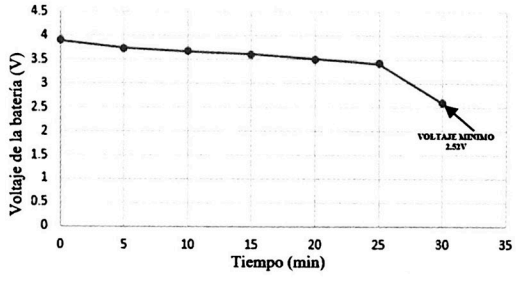


Figura 14. *Grafica de la descarga de la celda li-ion de la marca ULTRAFIRE.*

De estas gráficas, se pueden visualizar que al momento que se descarga la batería la caída de voltaje tiende a ser una función lineal.

Se visualiza también, que los tiempos de descarga difieren dependiendo de la capacidad de descarga de cada celda.

5. BMS (Sistema de gestión de batería)

Las baterías de iones de litio son sensibles a que funcionen fuera de su área de operación recomendada, lo que podría conducir a un menor tiempo de vida, daños y riesgo de explosión. Se aplica por ello un sistema de gestión de batería (BMS) para controlar y proteger la batería de las condiciones anormales [8]. En los últimos años, se ha prestado atención a esta tecnología de gestión de batería y, con el esfuerzo de los investigadores a lo largo del tiempo, su función ahora se puede definir explícitamente [9]:
A. Monitoreo en tiempo real de los estados de la batería. Al medir los parámetros característicos externos (como el voltaje externo, la corriente, la temperatura de la celda, etc.), con el algoritmo apropiado, BMS podría realizar la estimación y el monitoreo de los parámetros y estados internos de la batería como capacidad, estado de la carga, etc.
B. Uso eficiente de la energía de la batería.
C. Evitar la sobrecarga o sobre descarga de la batería.
D. Garantiza la seguridad del usuario y extiende la vida útil de la batería.

6. Materiales y resultados para la batería de 12.6V

6.1. Materiales

Tabla 4. Equipo usado para el circuito de la batería 12.6V

Nro	Materiales usados	Especificación
1	12 baterías de Li-Ion	De 3.7V
2	BMS	Des 3S y 20Amp
3	12 sockets para cada batería	Para una sola batería
4	Booster Step Up	De 3V a 30V
5	Fuente de 5V con indicador digital de voltaje y corriente	De 5Amp
6	Voltímetro digital	De 0v hasta 99.9V
7	Indicador de carga de batería	Para 3S

Se armó el circuito con celdas Li-ion y el BMS de la figura 16:

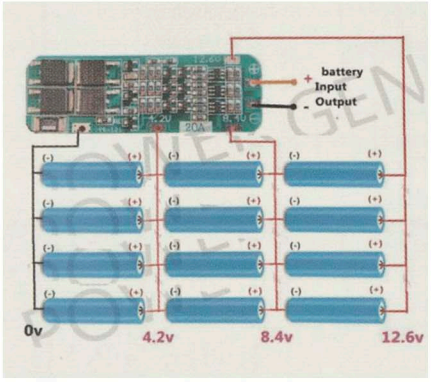


Figura 16. *Circuito de baterías 3S 4P conectado al BMS.*

El circuito del cargador usado es mostrado en la figura 16. Se utiliza una Fuente de 5V de 5Amperios que se conecta a un Booster Step Up. Este eleva el voltaje de 5v a 12.73v.

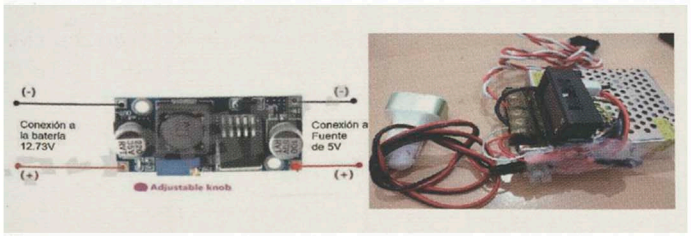


Figura 17. *Circuito del cargador de la batería de 12.6V.*

El sistema batería y cargador que se fabricó para este proyecto se muestra en la figura 18.

Figura 18. *Batería de Li-ion de 12.6V y el cargador.*

6.2. Resultados

Las mediciones obtenidas con el sistema de batería + cargador de la figura 18 se muestran en la tabla 4. El indicador digital mostró los voltajes obtenidos. Al momento de la carga y de la descarga de la batería.

Tabla 4. Valores del voltaje máximo y mínimo de una batería li-ion de 12.6V

Batería de 12.6V de iones de litio	
Voltaje máximo de carga y de corte	12.9V
Voltaje mínimo en la descarga y de corte	8.1V

7. Conclusiones

Al haberse realizado diversas pruebas con tres diferentes marcas de celdas li-ion se puede concluir que el voltaje máximo que se le da a la batería al momento de cargarlo y el voltaje mínimo de la batería al momento de ser descargada medidos del B+ y B- del TP4056 se encuentra entre los siguientes valores mostrados en la siguiente tabla.

Tabla 5. Valores del voltaje máximo y mínimo de una batería li-ion tomado del B+ y B- del TP4056

Celda de iones de litio	
Voltaje máximo de carga y de corte	4.24V-4.25V
Voltaje mínimo en la descarga y de corte	2.55V-2.60V

Si en la carga o descarga no se respetan estos valores dados en la tabla de arriba, la batería puede sobrecalentarse y explotar debido a una sobrecarga; o se podría auto descargar demasiado lo cual dejaría a la batería con una vida útil disminuida, pero si se carga o descarga a la batería como en la figura 10 y figura 13, respectivamente, se le está dando una mayor vida útil a la batería li-ion. Cuando la celda li-ion se descarga y llega a su voltaje mínimo, el TP4056 realiza un corte que hace que ya no disminuya el voltaje, y se puede ver que después del corte la batería infla el voltaje; es decir, sube un poco el voltaje de la batería, la batería se auto recarga.
Cuando se colocó una resistencia de 3.3Ohms y 10W, el TP4056 no permitía que ciertas celdas li-ion se descarguen, de esto se puede decir que, si se pone una resistencia de menor valor, el TP4056 en ese instante hará un corte, no dejará que la batería se descargue con una cantidad de corriente mayor a la que puede dar.
La capacidad de cada batería influye mucho en el tiempo que se demora cargar y descargar, a la vez con una batería de mayor capacidad se pueden usar resistencias de menor valor para descargarlos más rápido sin que el TP4056 lo corte en ese instante como se mencionó.

El OUT+ y el B+ del TP4056 están unidos.

En el momento que no se descarga la batería el B+ del TP4056 y el voltaje de la batería coinciden. Pero cuando empieza la descarga, el B+ del TP4056 y el voltaje de la celda li-ion no coinciden. Esto se debe al circuito interno

que hay en el TP4056.

La carga de la batería de 12.6V demoró un tiempo considerable de unas 8horas. Para bajar este tiempo, se hace necesario usar booster Step Up y una fuente de mayor amperaje. Se puede usar hasta 10 Amperios.

Agradecimiento

Al Centro de Tecnologías de Información y Comunicaciones (CTIC) de la UNI por el apoyo en el uso del laboratorio del tercer piso para poder desarrollar este trabajo de investigación.

1. J. Kim, J. Shin, C. Chun, and B. Cho, "Stable configuration of a li-ion series battery pack based on a screening process for improved voltage/ soc balancing" IEEE Transactions on Power Electronics, vol. 27, no. 1, pp. 411−424, 2012.

2. Tae-Ho Eom, Min-Ho Shin, Jun-Mo Kim, Jeong Lee, Chung-Yuen Won. Improved Charge Control Algorithm considering Temperature of Li-ion Battery. Sungkyunkwan University. Republic of Korea. 2017.

3. Bijani S., Laminas de Cu2O. Aplicación como electrodo, Tesis Doctoral, Universidad de Málaga. 2007.

4. Blog Baterías de litio. https://www.bateriasdelitio.net/?p=6. 2019.

5. Ignacio Martil https://blogs.cdecomunicacion.es/ignacio/2019/01/11/como-son-las-baterias-de-ion-litio/. 2019.

6. Casana N., Gomez P. Baterías de litio, Investigación y Ciencia. Barcelona. 1996.

7. NanJing Top Power ASIC Corp. TP4056 1A Standalone Linear Li-Ion Battery Charger with Thermal Regulation in SOP-8.

8. Waraporn Puviwatnangkurn, Bundit Tanboonjit, Nisai H. Fuengwarodsakul. Overcurrent Protection Scheme of BMS for Li-Ion Battery used in Electric Bicycles. Thailandia. 2013.

9. Jiuchun Jiang and Caiping Zhang. Fundamentals and applications of lithium -ion batteries in electric drive vehicles. John Wiley & Sons Singapore Pte Ltd. 2015.

Reglas para la Preparación de Artículos para la Revista REVCIUNI

En la revista REVCIUNI se publican artículos de investigación actual y divulgación científica, básica o aplicada, en las áreas de Física, Matemática, Química, Ciencia de la Computación, y afines. Los artículos se reciben en el Instituto de Investigación de la Facultad de Ciencias de la UNI.

Los artículos deben de ser originales, inéditos, que no se hayan publicado previamente ni se encuentren bajo consideración para ser publicados en otras revistas. Los artículos no deben presentar conclusiones conocidas, triviales, obvias y/o sin fundamento.

Los artículos serán recibidos por el Comité Científico el cual los enviará a uno o más árbitros para su revisión. El Comité Científico comunicará a los autores que sometieron el artículo la decisión de publicación así como las observaciones de los árbitros. Todos los artículos serán tratados de forma conficencial hasta su publicación.

Los artículos deben de ser escritos preferentemente en LaTex. La redacción y el formato del artículo deben seguir las siguientes indicaciones:

- El tipo de letra es normal Roman o equivalentes.
- Los márgenes son: de los lados derecho e izquierdo 1,5 cm y de arriba y abajo 2 cm.
- El título debe de estar centrado y escrito con letra normal de tamaño 14pt y en negrita.
- Debajo del título deben de ir los nombres completos de los autores con letra normal de tamaño 10pt. Después del nombre de cada autor, deben indicarse el lugar de trabajo y el correo electrónico de correspondencia (de uno de los autores) con letra cursiva y tamaño 10pt.
- El resumen debe escribirse en inglés y/o español, con letra normal tamaño 9pt. con un ancho del texto de 16,2 cm. debe contener entre 50 y 150 palabras e indicar al final las palabras claves. Primero va el resumen en el idioma en que se redactó el artículo.
- El texto se escribe con letra normal tamaño 10pt. En dos columnas separadas en 0,7 cm. Y debe ser dividido en secciones numeradas con números arábigos. El nombre de las secciones debe ser escrita en negrita tamaño 12pt. y centradas. Las subsecciones con letra negrita y centradas. Se recomienda que los artículos contengan las siguientes secciones: Introducción, Conclusiones y Agradecimientos (esta última no se numera).
- Al final va la sección sin numerar designada como Apéndice: Nombre del apéndice, en caso de haber varios apéndices van en secciones designadas como Apéndice A, Apéndice B, etc.
- Las fórmulas deben ser numeradas con números arábigos entre paréntesis en la margen derecha. La referencia de las fórmulas en el texto debe de haberse colocado entre paréntesis su número correspondiente.
- Toda letra latina que se utiliza en las fórmulas debe estar escrita en cursiva.
- Las funciones seno, coseno, logaritmo natural, y otras en esta categoría, se escriben sen, cos, ln, etc.
- Las tablas y figuras se enumeran con números arábigos. En la parte inferior de la tabla (figura), se colocará: Tabla (Figura) seguido del número correspondiente y un punto con letra negrita. La leyenda debe escribirse con letra cursiva, tamaño 10pt.
- Las citaciones del texto se hacen colocando el número correspondiente de la lista de referencias entre corchetes.
- Lista de referencias
  - Las referencias que se citan en el artículo es con número arábigos, en el orden de citación y va al final del artículo debajo de una línea horizontal, en dos columnas separadas en 0,7 cm. El tamaño de las letras es de 9pt.
  - Cada entrada en la lista de referencias debe estar citada en el texto.
  - Las comunicaciones personales se citan en el texto, pero no se incluyen en la lista de referencias.
  - Apellidos primero, seguidos de las iniciales del nombre.
  - Se utiliza el signo & antes del último autor. En español, se acepta la *y* en vez de &. En inglés, se acepta la *and* en vez de &.
  - En el caso de que la obra no tenga un autor, se coloca primero el título de la obra y luego la fecha.
  - Después de los nombres de los autores se coloca el nombre de la revista o libro con indicación al volumen, páginas y año entre paréntesis.

    Frittelli S., Kozameh C. and Newman E. T., J Matth. Phys. 36, 4975 (1995).
  - Cuando la referencia es a un capítulo de un libro editado, se escribe el nombre del editor, precedido por la abreviatura Ed.

    Arnold V. I., Mathematical Methods of Classical Mechanics, Ed. Springer, Berlín (1980).

Los artículos serán presentados previamente para su revisión en formato Portable Document Format (PDF) al e-mail: **postgradofc@uni.edu.pe.**

Revista de la Facultad de Ciencias de la UNI - REVCIUNI

Volumen 23, Número 1, enero-diciembre 2020

ISSN: 1813-3894

Publicada por el Instituto de Investigación de la Facultad de Ciencias
de la Universidad Nacional de Ingeniería

CONTENIDO