UNAH
UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

Maestría en Matemática

BOLETÍN DIVULGATIVO

OCTUBRE 2023

TEMÁTICA:

- CRIPTOGRAFÍA
- FLUJO VEHICULAR
- CURVAS ELÍPTICAS
- MODELOS MULTINIVEL
- P.D.E.
- MODELOS LINEALES DINÁMICOS

Presentación

Este documento fué desarrollado por la Coordinación General de Posgrados de la UNAH, presenta artículos divulgativos y de investigación desarrollados por profesionales del Seminario de Investigación de Ingeniería Matemática y de Estadística Matemática de la cuarta promoción del programa, cursos desarrollado durante el segundo período académico del año 2023. Se abarca una temática bastante amplia: criptogradía, flujo vehícular, curvas elípticas, modelos multinivel, ecuaciones diferenciales parciales (PDE) y modelos lineales dinámicos; en algunos de los trabajos se desarrolló una revisión bibliográfica de trabajos pertinentes y se resumió según lo comprendido por cada autor, en otros casos, se realizó avances en sus trabajos de tesis que incluso incluyen experimentación y por último, también se incluye algunas investigaciones originales.

El objetivo principal de desarrollar este documento es que a futuro, en base a la experiencia obtenida y después de tener varias experiencias similares, se transforme en una revista científica de Matemáticas, cuestión que requiere de mucho trabajo por parte del equipo de profesores investigadores del programa y otros colaboradores externos; además de ser una muestra de que en el programa de maestría en Matemáticas y por parte de la Coordinación General de Posgrados de la Facultad de Ciencias, se está desarrollando en los estudiantes un espíritu investigador.

Todas las revisiones bibliográficas y temas aquí presentados se encasillan dentro de las líneas de investigación de la UNAH, entre los temas prioritarios abarcados se encuentran: ciencia, cambio climático y vulnerabilidad, productividad, infraestructura y desarrollo territorial. Esto evidencia que la Coordinación General de Posgrados de la Facultad de Ciencias está sumamente interesada en colaborar con las prioridades investigativas de la universidad y mantiene un compromiso con vincularse con la

sociedad.

En esta ocasión también se presenta una publicación original realizada por el matemático hondureño José Mauricio Alvarenga, Licenciado en Matemática de la UNAH, Máster en Matemática con orientación en Ingeniería Matemática de la UNAH, entre otros estudios e investigaciones realizadas. Actualmente es Profesor de matemáticas de la Carrera de Matemática y de la Maestría en Matemática de la UNAH. Esta publicación le agradece enormemente este aporte, ya que contribuye al crecimiento de la misma y es un paso importante hacia la realización del objetivo de transformarse en una revista de matemáticas.

Octubre del año 2023, Ciudad Universitaria

Tegucigalpa, M.D.C., Honduras

® Coordinación General de Posgrados de la Facultad de Ciencias
Maestría en Matemáticas - UNAH
Edificio F1, Segundo Piso, Ciudad Universitaria
Tegucigalpa, M.D.C. Honduras.
https://mm.unah.edu.hn/
maestria.matematica@unah.edu.hn
Tel. 2216-3000 Ext. 100647

Contenido

Un modelo de flujo de tráfico vehicular en una red vial

Jose Alvarenga[a], Ivan Henriquez[b]

Tegucigalpa, Honduras

[a] *jose.alvarenga@unah.edu.hn*
[b] *ivan.enriquez@unah.edu.hn*

Abstract

Desde la década de los cuarenta se han desarrollado diferentes enfoques teóricos y prácticos para entender la dinámica del tráfico vehicular, todos estos esfuerzos se ven reunidos en una teoría conocida con el nombre de modelos continuos del tráfico vehicular. Este trabajo se encuentra en el contexto de los *modelos macroscópicos*; una de las principales ramas de los modelos continuos en donde el tráfico se entiende como un fluido gobernado por una ecuación de transporte. En este artículo se estudio un modelo macroscópico para una red de carreteras basado en el trabajo [5], integrando una función de flujo que depende de la variable espacial.

Por otro lado se desarrollo un marco computacional para resolver el sistema de ecuaciones diferenciales inherente al modelo.

Keywords: Red vial, flujo vehicular, modelos macróscopicos, dependencia espacial, condiciones de flujo, optimización en la frontera, problema de Riemann.

1. Introducción

Cuando se asume que algunas de las cantidades asociadas al tráfico vehicular (e.g. velocidad del tráfico) dependen continuamente del tiempo, entonces se dice que el modelo que se está estudiando es continuo.

Todos los modelos continuos de tráfico vehicular tienen un origen común en el trabajo realizado por Greenshields en [1], donde se midieron las velocidades promedio de un conjunto de vehículos, sometidos a diferentes escenarios. Como una de las conclusiones importantes de este trabajo se extrajo que la velocidad promedio depende de las distancias entre los vehículos del sistema. La forma precisa en la que esta relación se sostiene, es un aspecto que está lejos de ser trivial, este es uno de los puntos que hace que en la actualidad se disponga de un amplio abanico de formas en las que estas dos cantidades están relacionadas.

Un último aspecto define a los modelos continuos del tráfico vehicular, y esto recae explícitamente en la palabra *continuo*, además de que en principio se asume que diferentes variables del tráfico se encuentran relacionadas, también se presume que es posible predecir la evolución de algunas de las variables del tráfico en función del tiempo, esta última característica es la que da lugar a dos ramas sobresalientes de esta teoría; los modelos macroscópicos y microscópicos. En los modelos macroscópicos se entiende al tráfico

como un flujo dinámico controlado por una ecuación de transporte con flujo no lineal y esta será la óptica desde la que se desarrollará este trabajo.

2. Descripción del modelo

La visión macroscópica del tráfico vehicular proviene de la mecánica de fluidos y se le puede atribuir a Michael James Lighthill y Gerald Whitham, su trabajo se puede encontrar plasmado en los artículos [2, 3].

Además de las dos publicaciones de Whitham y Lighthill, de manera independiente, Richards desarrolla en [4] un modelo del tráfico desde la óptica, nuevamente, de los fluidos mecánicos; en este trabajo se modela al tráfico como un fluido compresible (la densidad no necesariamente se mantiene constante) donde la velocidad del fluido depende únicamente de la densidad, es por esto que a este tipo de modelos se les conoció en sus inicios como los modelos LWR(Lighthill-Whitham-Richards).

Uno de los principales atractivos del enfoque macroscópico es su capacidad para representar características complejas del tráfico con una formulación compacta y sencilla, esto hace que este modelo tenga preferencia ante situaciones en las que no se requiere una resolución alta de la dinámica del tráfico vehicular, como es el caso de redes de tráfico vehicular.

En términos precisos la dinámica del tráfico en el enfoque macroscópico viene dada por la siguiente ecuación diferencial:

$$q_t(x,t) + [f(x,t)]_x = 0, \tag{1}$$

donde $q(x,t)$ representa la densidad vehicular promedio (número de vehículos por unidad de longitud) y $f(x,t)$ es el flujo vehicular (número de carros por unidad de tiempo) en el tiempo t a la altura del punto x. Además $f(x,t) = q(x,t)V(x,t)$, donde a la función $V(x,t)$ se le conoce como campo de velocidades.

Formalmente el modelo matemático que se estudió, se encuentra descrito en el trabajo [1], donde el flujo se relaciona con la densidad por medio de una expresión parabólica, específicamente esta relación se puede escribir de la siguiente forma:

$$f(x,t) \equiv f(q(x,t)) = v_{max} q(x,t)\left(1 - \frac{q(x,t)}{q_{max}}\right), \tag{2}$$

donde q_{max} y v_{max} representan respectivamente la densidad y la velocidad máxima. Aquí se observa que el campo de velocidades viene representado en la ecuación (2) por $V(x,t) = v_{max}\left(1 - \frac{q(x,t)}{q_{max}}\right)$.

Una de las mejoras que se introducen en este trabajo, en comparación con el modelo expuesto en [1], es la posibilidad de captar la sensibilidad del flujo vehicular a la variable espacial, esta característica se puede conseguir haciendo variar el parámetro de la velocidad máxima; se introduce entonces, la variación de la velocidad máxima, a través de la función $v : [0, L] \longrightarrow \mathbb{R}^+$, donde L es la longitud de la carretera de estudio. De esta forma se define el flujo que depende de la variable espacial:

$$f(q,x) = v(x) q\left(1 - \frac{q}{q_{max}}\right). \tag{3}$$

En este artículo se considerará una dependencia de la siguiente forma:

$$v(x) = ax + b \tag{4}$$

La ecuación (4) supone dos ventajas, una es que la dependencia no aumenta demasiado la complejidad, al ser lineal en la variable espacial, otra es que cumple con la funcionalidad de captar la sensibilidad del flujo en aspectos propios de la carretera de estudio y no de la densidad únicamente.

Con lo anterior la dinámica del tráfico vendría dada por la siguiente ecuación diferencial:

$$q_t(x,t) + \left[(ax+b)q(x,t)\left(1 - \frac{q(x,t)}{q_{max}}\right)\right]_x = 0. \tag{5}$$

En lo que sigue de esta sección se definirá la dinámica del modelo matemático del tráfico para un sistema de carreteras, para ello se uso uno de los primeros trabajos en esta dirección. Holden y Risebro en [5] proponen un modelo de esta clase donde un conjunto de carreteras se visualiza como un grafo dirigido, ver Figura 1.

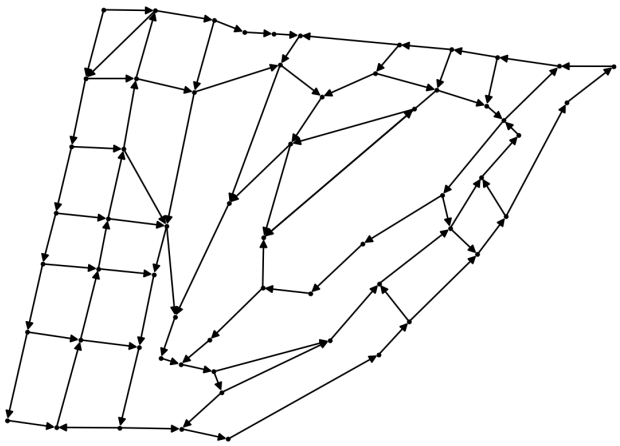


Figura 1: Grafo dirigido.

Definición 1 (Red de tráfico vehicular)**.** *Supóngase que en un sistema de calles se tienen N carreteras y n intersecciones, un grafo de carreteras es un grafo conexo $G = (V, A)$, donde el conjunto de aristas se asocia con las carreteras y los vértices son las intersecciones de estas.*

En la Definición 1 el grafo G contendrá los siguientes atributos:

1. Sea $l \in A$, entonces se denotará la densidad en esta carretera como $q_l(x,t)$.
2. Para cada arista $l \in A$ se tendrán los atributos q^l_{max}, a_l, b_l.

3. Para cada arista $l \in A$ la densidad $q_l(x,t)$ vendrá dada por la ecuación (5), es decir:

$$[q_l(x,t)]_t + [f_l(q_l(x,t),x)]_x = 0.$$

Donde $f_l(q_l(x,t),x) = (a_l x + b_l)q_l(x,t)\left(1 - \frac{q_l(x,t)}{q^l_{max}}\right)$.

4. Para cada arista $l \in A$ se denotará a la longitud de la carretera como L_l, con esto la función de densidad quedará definida para $x \in [0, L_l]$.
5. Además de considerar el modelo de [5], en trabajos posteriores a este, se considero incluir la intencionalidad del flujo en las intersecciones, como ejemplo considérese el trabajo [6]. Formalmente, para cada $i \in V$ sean C^e_i el conjunto de aristas de entrada y C^s_i el conjunto de aristas de salida del vértice i, entonces se asumirá que para cada $l \in C^e_i$ y cada $k \in C^s_i$ existirán unos coeficientes no negativos α^i_{lk} tales que:
   (a) Para cada $l \in C^e_i$ se tiene que $\sum_{k \in C^s_i} \alpha^i_{lk} = 1$.
   (b) Para cada $k \in C^s_i$ se tiene que $f_k(q_k(0,t),0) = \sum_{l \in C^e_i} \alpha^i_{lk} f_l(q(L_l,t),L_l)$.
6. Con la consideración anterior se cumple que para todo $i \in V$:

$$\sum_{k \in C^s_i} f_k(q_k(0,t),0) = \sum_{l \in C^s_i} f_l(q_l(L_l,t),L_l). \tag{6}$$

   A la igualdad (6) se le conoce como condición de *Rankine-Hugoniot*, la cual expresa que el flujo en las intersecciones se deben conservar.
7. Sea $i \in V$ tal que $grad(i) = 1$ (solo existe una arista que tiene a este vértice como uno de sus elementos), entonces en este vértice no se impone la condición de *Rankine-Hugoniot*.
8. Para $i \in V$ denótese $grade(i)$ como el número de aristas de entrada al vértice i y $grads(i)$ como el número de aristas de salida del vértice i.

Con todo lo anterior se define formalmente el modelo matemático de estudio para este trabajo.

Definición 2 (Modelo matemático)**.** *Para cada $k \in A$ se tiene la ecuación diferencial:*

$$[q_k(x,t)]_t + [f_k(q_k(x,t),x)]_x = 0$$

Sujeto a las siguientes condiciones:

- *Condición inicial :* $q_k(x,0) = g_k(x)$.

Para cada $i \in V$:

- *Condición de frontera: Si* $grade(i) = 1$ *y* $grads(i) = 0$ *entonces* $q_s(L_s,t) = h_s(t)$ *donde* $s \in C^e_i$.
- *Condición de frontera: Si* $grade(i) = 0$ *y* $grads(i) = 1$ *entonces* $q_s(0,t) = h_s(t)$ *donde* $s \in C^s_i$.
- *Condición de flujo: Si* $grade(i) + grads(i) \geq 2$,
  entonces para cada $k \in C^s_i$ *se tiene que* $f_k(q_k(0,t),0) = \sum_{l \in C^e_i} \alpha^i_{lk} f_l(q(L_l,t),L_l)$.

Donde las funciones g_k y h_s son dadas.

3. Análisis de la condición de flujo

En esta sección se hará un análisis que servirá para construir el marco computacional para el problema dado en la Definición 2.

3.1. Condiciones de frontera

Aquí se analizarán los requisitos suficientes para que una condición de frontera constante defina una solución consistente.

Se inicia con el siguiente resultado, el cual establece la solución implícita de un problema de valor en la frontera con condición constante.

Proposición 1. *Considérese el problema $q_t + [f(q,x)]_x = 0$ con condición de frontera, $q(\overline{x},t) = \overline{q}$, donde $\overline{x}$ y $\overline{q}$ son constantes. Entonces la solución a este problema viene dada implícitamente por la ecuación*

$$f(q,x) = f(\overline{q},\overline{x})$$

donde $f(q,x) = (ax+b)q\left(1 - \dfrac{q}{q_{max}}\right)$.

La siguiente proposición establece las condiciones suficientes para que la solución que se describe en la Proposición 1 este bien definida para cualquier x.

Proposición 2. *Suponga las mismas hipótesis que en la Proposición 1. Además defínanse las constantes $v_{max} = \max_x v(x)$ y $v_{min} = \min_x v(x)$, entonces el problema tiene al menos una solución, sí:*

$$v_{min} g(q_{opt}) \geq v_{max} g(\overline{q}),$$

donde $q_{opt} = q_{max}/2$ *y* $g(q) = q\left(1 - \dfrac{q}{q_{max}}\right)$.

Demostración. Se define inicialmente el polinomio:

$$r(q) = f(q,x) - f(\overline{q},\overline{x}) = v(x)g(q) - v(\overline{x})g(\overline{q}).$$

En la definición anterior $r(q)$ es un polinomio cuadrático cóncavo hacia abajo, además nótese que:

$$-f(\overline{q},\overline{x}) = r(0) = r(q_{max}).$$

De lo anterior se tiene que el vértice de este polinomio se encuentra entre 0 y q_{max}, no es difícil probar que esto sucede en $q = \dfrac{q_{max}}{2}$. Con el siguiente razonamiento:

$$\begin{aligned} r\left(\frac{q_{max}}{2}\right) =& v(x)g\left(\frac{q_{max}}{2}\right) - v(\overline{x})g(\overline{q}) \\ \geq& v_{min} g(q_{opt}) - v_{max} g(\overline{q}) \\ =& 0, \end{aligned}$$

se llega a que $r\left(\dfrac{q_{max}}{2}\right) \geq 0$. De esta forma se puede encontrar para cualquier x al menos un q que verifique la ecuación $r(q) = 0$, que es lo que se quería probar. □

La proposición anterior se puede reescribir, afirmando que la siguiente condición es suficiente para que la solución dada en la Proposición 1 esté bien definida.

$$\overline{q} \in [0, q_{opt}(1-\beta)] \cup [q_{opt}(1+\beta), q_{max}] \tag{7}$$

donde $\beta^2 = 1 - \dfrac{v_{min}}{v_{max}}$.

3.2. Problema de Riemann en la frontera

En este apartado se analizará la solución de la ecuación diferencial (5) cuando se tiene una condición de frontera y una condición inicial constante. Se describe a continuación el problema:

$$\begin{aligned} q_t + \left[(ax+b)q\left(1 - \frac{q}{q_{max}}\right)\right]_x &= 0, \\ q(x, t_0) &= q_0, \\ q(\overline{x}, t) &= q_1. \end{aligned}$$

La solución al problema con valor inicial del planteamiento anterior es:

$$q(x,t) = Q_0(t) \equiv \frac{q_0 q_{max}}{(q_{max} - q_0)e^{a(t-t_0)} + q_0}. \tag{8}$$

Para la solución en la frontera se tiene que:

$$q(x,t) = Q_1(x) \equiv q_{opt} \pm \sqrt{\alpha(x)[q_{opt} - q_1]^2 + \overline{\alpha}(x) q_{opt}^2}, \tag{9}$$

donde $q_{opt} = \dfrac{q_{max}}{2}$, $g(q) = q\left(1 - \dfrac{q}{q_{max}}\right)$, $\alpha(x) + \overline{\alpha}(x) = 1$ y $\alpha(x) = \dfrac{a\overline{x} + b}{ax + b}$.
Para que la solución sea continua, se sigue la siguiente regla:

$$Q_1(x) = \begin{cases} q_{opt} + \sqrt{\alpha(x)[q_{opt} - q_1]^2 + \overline{\alpha}(x))q_{opt}^2} & \text{si } q_1 > q_{opt} \\ \\ q_{opt} - \sqrt{\alpha(x)[q_{opt} - q_1]^2 + \overline{\alpha}(x))q_{opt}^2} & \text{si } q_1 < q_{opt} \end{cases}$$

De las relaciones (8) y (9) se puede observar que las curvas características provocadas por la condición inicial, son lineas rectas horizontales, mientras que las curvas características producidas por la condición de frontera, son lineas rectas verticales; a diferencia de los modelos donde el flujo solo depende de la densidad, aquí siempre se presentan este tipo de curvas características que a su vez producen el fenómeno conocido como *choque de curvas características*. El siguiente resultado describe la naturaleza de este choque.

Proposición 3. *Considere el problema dado en la ecuación* (5) *con condición inicial* $q(x,t_0) = q_0$ *y condición de frontera* $q(\overline{x},t) = q_1$ *para todo* $t \geq t_0$. *Sea* $(x(t),t)$ *una curva que divide el plano* (x,t) *de manera que la solución a un lado corresponde con la generada por el problema de valor inicial* $q(x,t_0) = q_0$, *la cual viene dada por la ecuación*(8) *y del lado restante, la solución generada por el problema de valor en la frontera* $q(\overline{x},t) = q_1$, *la cual viene dada por la ecuación* (9).
Entonces, para que la ley de conservación se mantenga, la curva debe satisfacer la siguiente ecuación diferencial:

$$x'(t) = \frac{(ax(t)+b)(1-Q_1(x(t))-Q_0(t))}{q_{max}}$$

con condición inicial $x(t_0) = \overline{x}$.
Finalmente, si $x(t)$ *cumple la ecuación diferencial que se mencionó antes, entonces la solución al problema de Riemman viene dada por la siguiente igualdad:*

$$q(x,t) = \begin{cases} Q_0(t) & si\ x < x(t) \\ Q_1(x) & si\ x \geq x(t) \end{cases}.$$

Demostración. Como $q(x,t)$ verifica la ecuación de conservación integral, entonces se verifica que:

$$\frac{d}{dt}\int_a^b q(x,t)dx = f(q(a,t),a) - f((b,t),b). \tag{10}$$

Por otro lado se tiene que:

$$\begin{aligned}
&\frac{d}{dt}\int_a^b q(x,t)dx = \frac{d}{dt}\int_a^{x(t)} q(x,t)dx + \frac{d}{dt}\int_{x(t)}^b q(x,t)dx \\
&= \int_a^{x(t)} q_t(x,t)dx + \int_{x(t)}^b q_t(x,t)dx + Q_0(t)x'(t) - Q_1(x(t))'(t) \\
&= -\int_a^{x(t)} [f(q(x,t),x)]_x dx - \int_{x(t)}^b [f(q(x,t),x)]_x dx + (Q_0(t) - Q_1(x(t)))x'(t) \\
&= -f(Q_0(t),x(t)) + f(q(a,t),a) - f(q(b,t),b) + f(Q_1(x(t)),x(t)) + (Q_0(t) - Q_1(x(t)))x'(t),
\end{aligned}$$

entonces se tiene la siguiente relación:

$$\frac{d}{dt}\int_a^b q(x,t)dx = -f(Q_0(t),x(t)) + f(q(a,t),a) - f(q(b,t),b) + f(Q_1(x(t)),x(t)) + (Q_0(t) - Q_1(x(t)))x'(t). \tag{11}$$

Comparando las ecuaciones 10 y 11 se llega a:

$$0 = -f(Q_0(t),x(t)) + f(Q_1(x(t)),x(t)) + (Q_0(t) - Q_1(x(t)))x'(t),$$

de esta última expresión se puede deducir el siguiente razonamiento:

$$\begin{aligned} x'(t) =& \frac{f(Q_1(x(t)), x(t)) - f(Q_0(t), x(t))}{Q_1(x(t)) - Q_0(t)} \\ =& \frac{(ax(t)+b)Q_1(x(t))(1 - Q_1(x(t))/q_{max}) - (ax(t)+b)Q_0(t)(1 - Q_0(t)/q_{max})}{Q_1(x(t)) - Q_0(t)} \\ =& \frac{(ax(t)+b)(1 - Q_1(x(t)) - Q_0(t))}{q_{max}}. \end{aligned}$$

Esto último es lo que se quería probar. □

Ahora se expondrán dos resultados que se usarán para poder determinar un dominio para las densidades en la frontera, de manera que estas se impongan sobre la condición inicial, en cierto sentido.

Para empezar supóngase que se tiene el problema (5) con condiciones $q(x,0) = q_0$ para todo $x < \overline{x}$ y $q(t,\overline{x}) = q_1$ para todo $t > t_0$. Sea $x(t)$ la curva definida en la Proposición 3, las condiciones siguientes son suficientes para garantizar que $x'(t) < 0$ para algún intervalo de tiempo de la forma $[t_0, T]$ y que se verifique la Proposición 2:

- Si $q_0 \leq q_{opt}$ entonces $q_1 \in [\max\{\overline{q}_0, q_{opt}(1+\beta)\}, q_{max}]$.
- Si $q_{opt} < q_0 \leq q_{opt}(1+\beta)$ entonces $q_1 \in [q_{opt}(1+\beta), q_{max}]$.
- Si $q_0 > q_{opt}(1+\beta)$ entonces $q_1 \in [\overline{q}_0, q_{opt}(1-\beta)] \cup [q_{opt}(1+\beta), q_{max}]$.

Donde $\overline{q_0} = q_{max} - q_0$. Una condición suficiente para que se cumplan las condiciones anteriores se expresa a continuación:

$$q_1 \in [\max\{\overline{q}_0, q_{opt}(1+\beta)\}, q_{max}]. \tag{12}$$

De forma similar, supóngase que se tiene el problema (5) con condiciones $q(x,0) = q_0$ para todo $x > \overline{x}$ y $q(t,\overline{x}) = q_1$ para todo $t > t_0$. Sea $x(t)$ la curva definida en la Proposición 3, las condiciones siguientes son suficientes para garantizar que $x'(t) > 0$ para algún intervalo de tiempo de la forma $[t_0, T]$ y que se verifique la Proposición 2:

- Si $q_0 \geq q_{opt}$ entonces $q_1 \in [0, \min\{\overline{q}_0, q_{opt}(1-\beta)\}]$.
- Si $q_{opt}(1-\beta) \leq q_0 < q_{opt}$ entonces $q_1 \in [0, q_{opt}(1-\beta)]$.
- Si $q_0 < q_{opt}(1-\beta)$ entonces $q_1 \in [q_{opt}(1+\beta), \overline{q}_0] \cup [0, q_{opt}(1-\beta)]$.

Al igual que antes, una condición suficiente para que se cumplan las condiciones anteriores se expresa a continuación:

$$q_1 \in [0, \min\{\overline{q}_0, q_{opt}(1-\beta)\}]. \tag{13}$$

Estas últimas observaciones son parte de las bases para la construcción de un marco computacional para el problema planteado en la Definición 2.

4. Marco computacional

En esta sección se desarrollarán un conjunto de herramientas numéricas que resolverán el problema dado en la Definición 2 usando los resultados de las secciones anteriores.

4.1. Método de volúmenes finitos

El problema que se quiere abordar numéricamente aquí, es el dado en la Definición 5 sujeto a las siguientes condiciones:

$$\begin{aligned} &(x,t) \in [0,L] \times [0,T], && q(0,t) = q_1(t), \\ &q(x,0) = q_0(x), && q(L,t) = q_2(t). \end{aligned}$$

Para resolver este problema se usará el método de volúmenes finitos. Dentro de este tipo de métodos numéricos es necesario determinar una función conocida como *flujo numérico*, en [7] se describen una clase de flujos numéricos conocidos con el apellido del matemático aplicado Sergei Godunov, en lo que resta de este apartado se construirá el flujo numérico de Godunov para este tipo de problemas. Además se encontrará la región de estabilidad para este esquema numérico y se validará el código implementado.

Para aplicar el método de volúmenes finitos se realiza en espacio y tiempo la partición siguiente:

$$\begin{aligned} &x_i = i\Delta x, && \text{para } i \in \{1,2,3,\cdots,n+1\}, \\ &t_j = j\Delta t, && \text{para } j \in \{1,2,3,\cdots,m\}, \end{aligned}$$

donde $x_0 = 0$, $x_{n+1} = L$, $t_0 = 0$ y $t_m = T$. Para implementar el método se necesitan ajustar los valores iniciales y de frontera de la siguiente forma:

$$\begin{aligned} Q_i^0 \approx & \frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} q_0(x)dx, \\ L_j \approx & \frac{1}{\Delta t} \int_{t_j}^{t_{j+1}} q_1(t)dt, \\ R_j \approx & \frac{1}{\Delta t} \int_{t_j}^{t_{j+1}} q_2(t)dt. \end{aligned}$$

Se resalta que el método de volúmenes finitos intenta aproximar los valores Q_i^j en el siguiente sentido:

$$Q_i^j \approx \frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} q(x,t_j)dx.$$

En los esquemas numéricos del método de volúmenes finitos se plantea que Q_i^j se puede aproximar de la siguiente forma:

$$Q_i^j = Q_i^j - \frac{\Delta t}{\Delta x}(F(Q_i^j, Q_{i+1}^j, x_i) - F(Q_{i-1}^j, Q_i^j, x_i)). \tag{14}$$

Donde F viene a ser el flujo numérico. El flujo numérico de Godunov se puede encontrar resolviendo la siguiente integral:

$$\frac{1}{\Delta t}\int_{t_j}^{t_{j+1}} f(q(x_i,t),x_i)dt.$$

Esta integral se resuelve con la información promedio alrededor del punto x_i, es decir se resuelve el problema en 4 sujeto a $q(x,t_j)=Q_{i-1}^j$ para $x<x_i$ y $q(x,t_j)=Q_i^j$ para $x>x_i$. Se puede probar que el aporte a la solución, proveniente de los valores Q_i^j y Q_{i-1}^j es:

$$q(x,t)=\frac{Q_{i-1}^j q_{max}}{\overline{Q}_{i-1}^j e^{a(t-t_j)}+Q_{i-1}^j}, \tag{15}$$

$$q(x,t)=\frac{Q_i^j q_{max}}{\overline{Q}_i^j e^{a(t-t_j)}+Q_i^j}. \tag{16}$$

Donde la $\overline{Q}_i^j=q_{max}-Q_i^j$. Con la información anterior se puede calcular el flujo numérico:

$$F(Q_{i-1}^j,Q_i^j,x_i)\equiv\frac{1}{\Delta t}\int_{t_j}^{t_{j+1}} f(q(x_i,t),x_i)dt=\frac{v(x_i)Q_k^j\overline{Q}_k^j(e^{a\Delta t}-1)}{a(\overline{Q}_k^j e^{a\Delta t}+Q_k^j)\Delta t}, \tag{17}$$

en la expresión anterior, k es una variable que puede tomar los valores de i o $i-1$; para poder determinar el valor correcto de k se necesitan resolver en la mayoría de casos, los choques que resultan de este problema. Si se sigue un razonamiento similar al expuesto en la Proposición 3 se llega a que la curva de choque, $x(t)$ es igual a:

$$x(t)=\frac{(ax_i+b)e^{-a(t-t_j)}(\overline{Q}_i^j e^{a(t-t_j)}+Q_i^j)(\overline{Q}_{i-1}^j e^{a(t-t_j)}+Q_{i-1}^j)-bq_{max}^2}{aq_{max}^2}. \tag{18}$$

La derivada de esta curva viene dada por la siguiente expresión:

$$x'(t)=\frac{(\overline{Q}_i^j\overline{Q}_{i-1}^j e^{2a(t-t_j)}-Q_{i-1}^j Q_i^j)e^{-a(t-t_j)}(ax_i+b)}{q_{max}^2}.$$

De forma precisa $k=i-1$ cuando $x'(t)>0$ durante un tiempo mayor a t_j y $k=i$ si $x'(t)<0$ durante algun mayor a t_j.

4.2. Análisis de estabilidad

Considérese el problema de calcular de calcular $F(Q_i^j,Q_{i+1}^j,x_i)-F(Q_{i-1}^j,Q_i^j,x_i)$ en la expresión (14), es posible que en el proceso del cálculo de $F(Q_i^j,Q_{i+1}^j,x_i)$ y $F(Q_{i-1}^j,Q_i^j,x_i)$ las curvas de choque, inherentes a cada valor, coincidan para algún $t=T$ entre t_j y t_{j+1}, por lo que el cálculo en (17) sería

incorrecto. Suponiendo que las dos curvas de choque coinciden (es posible que esto no suceda) y usando la forma exacta de las curvas de choque dada en (18) se calculará el tiempo $t = T$ en que estas se cruzan:

$$\begin{aligned}
\frac{v(x_i)e^{-a\Delta T}(\overline{Q}^j_{i-1}e^{a\Delta T}+Q^j_{i-1})(\overline{Q}^j_{i}e^{a\Delta T}+Q^j_{i})}{aq^2_{max}} =& \frac{v(x_{i+1})e^{-a\Delta T}(\overline{Q}_{i}e^{a\Delta T}+Q^j_{i})(\overline{Q}^j_{i+1}e^{a\Delta T}+Q^j_{i+1})}{aq^2_{max}},\\
v(x_i)e^{-a\Delta T}(\overline{Q}^j_{i-1}e^{a\Delta T}+Q^j_{i-1})(\overline{Q}^j_{i}e^{a\Delta T}+Q^j_{i}) =& v(x_{i+1})e^{-a\Delta T}(\overline{Q}_{i}e^{a\Delta T}+Q^j_{i})(\overline{Q}^j_{i+1}e^{a\Delta T}+Q^j_{i+1}),\\
v(x_i)(\overline{Q}^j_{i-1}e^{a\Delta T}+Q^j_{i-1}) =& v(x_{i+1})(\overline{Q}^j_{i+1}e^{a\Delta T}+Q^j_{i+1}),\\
v(x_i)(\overline{Q}^j_{i-1}e^{a\Delta T}+Q^j_{i-1}) =& (v(x_i)+a\Delta x)(\overline{Q}^j_{i+1}e^{a\Delta T}+Q^j_{i+1}),\\
e^{a\Delta T}[v(x_i)(Q^j_{i+1}-Q^j_{i-1})-a\Delta x\overline{Q}^j_{i+1}] =& v(x_i)(Q^j_{i+1}-Q^j_{i-1})+a\Delta xQ^j_{i+1},\\
\Delta T =& \frac{1}{a}\ln\left(\frac{v(x_i)(Q^j_{i+1}-Q^j_{i-1})+a\Delta xQ^j_{i+1}}{v(x_i)(Q^j_{i+1}-Q^j_{i-1})-a\Delta x\overline{Q}^j_{i+1}}\right),
\end{aligned}$$

donde $\Delta T = T - t_j$ y $v(x) = ax + b$. Por lo tanto una condición necesaria para que las curvas de choquen no se intercepten se expresa por la siguiente desigualdad:

$$\Delta t < \frac{1}{a}\ln\left(\frac{v(x_i)(Q^j_{i+1}-Q^j_{i-1})+a\Delta xQ^j_{i+1}}{v(x_i)(Q^j_{i+1}-Q^j_{i-1})-a\Delta x\overline{Q}^j_{i+1}}\right). \tag{19}$$

Para que exista la posibilidad de que las curvas coincidan, se debe verificar que:

$$Q^j_{i+1}-Q^j_{i-1} > \frac{\max\{a\overline{Q}^j_{i+1}, -aQ^j_{i+1}\}\Delta x}{v}. \tag{20}$$

La condición (19) resulta poco práctica dado que compromete los índices de la partición. Se harán algunas estimaciones para encontrar una condición suficiente que garantice tal condición.

Si (20) se cumple, entonces la siguiente desigualdad también se mantiene:

$$\frac{1}{a}\ln\left(\frac{v(x_i)(Q^j_{i+1}-Q^j_{i-1})+a\Delta xQ^j_{i+1}}{v(x_i)(Q^j_{i+1}-Q^j_{i-1})-a\Delta x\overline{Q}^j_{i+1}}\right) \geq \frac{1}{a}\ln\left(\frac{v(x_i)q_{max}+a\Delta xQ^j_{i+1}}{v(x_i)q_{max}-a\Delta x\overline{Q}^j_{i+1}}\right). \tag{21}$$

Además se puede probar que:

$$\frac{1}{a}\ln\left(\frac{v(x_i)q_{max}+a\Delta xQ^j_{i+1}}{v(x_i)q_{max}-a\Delta x\overline{Q}^j_{i+1}}\right) \geq \frac{1}{a}\ln\left(\frac{v_{max}+a\Delta x}{v_{max}}\right), \tag{22}$$

donde $v_{max} = \max_x v(x)$. Usando (22), (21) y (19) se llega a que una condición suficiente para que las curvas de choque no coincidan es:

$$\Delta t < \frac{1}{a}\ln\left(\frac{v_{max}+a\Delta x}{v_{max}}\right). \tag{23}$$

Proposición 4. *La siguiente expresión es una condición suficiente para que el cálculo del flujo numérico de Godunov* (17) *sea correcto:*

$$\Delta t < \frac{1}{a} \ln \left(\frac{v_{max} + a\Delta x}{v_{max}} \right),$$

donde $v_{max} = \max_x(ax + b)$.

4.3. Condiciones de acoplamiento

Con el método descrito al inicio de esta sección se puede resolver, en parte, el problema dado en la Definición 2 en cada una de las aristas del grafo de carreteras. Para completar el marco computacional, se explicará en este apartado, como intervienen las condiciones de frontera en el acoplamiento de estas carreteras.

Como se hizo al inicio de esta sección, se creará una partición tanto en tiempo como en espacio para cada una de las carreteras, con esto se seguirá la siguiente notación: Para el grafo $G = (V, A)$ que define al sistema de carreras, sea $l \in A$, la notación de los elementos de la partición incorporarán un subíndice derecho para los elementos promedios de la densidad, es decir:

$$_lQ_i^j \approx \frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} q_l(x, t_j)dx.$$

En la expresión anterior la partición en espacio podría ser diferente, sin embargo para evitar una notación cargada se decidió dejar la misma partición para todas las carreteras, por otro lado, la partición en tiempo si debe ser igual. En lo que resta de esta sección, aparecerán algunas constantes y variables con índices de carretera, para indicar su pertenencia a estas.

El principal objetivo de esta sección es determinar la información en la frontera, es decir, se desea conocer en cada instante de tiempo los valores en la frontera de cada carretera de manera que no violen la conservación de flujo en cada uno de los vértices. Para ser precisos se desean encontrar dos conjuntos de valores $\{R_j^l\}_{l\in A}$ y $\{L_j^l\}_{l\in A}$ de manera que R_j^l represente la información de lado derecho de la carretera l en el intervalo de tiempo de $[t_j, t_{j+1}]$ y que L_j^l represente la información de lado izquierdo de la carretera l en el intervalo de tiempo de $[t_j, t_{j+1}]$.

Sea $l = (u, v) \in A$. Si $grads(u) = 1$ y $grade(u) = 0$ entonces de acuerdo a la Definición 2:

$$L_j^l = \frac{1}{\Delta t} \int_{t_j}^{t_{j+1}} h_u(t)dt,$$

y si $grade(v) = 1$ y $grads(v) = 0$ entonces:

$$R_j^l = \frac{1}{\Delta t} \int_{t_j}^{t_{j+1}} h_v(t)dt.$$

Los dos casos anteriores se resuelven directamente de la definición del problema. Por otro lado sea $i \in V$ tal que $grad(i) > 1$, entonces usando la Definición 2 se debe cumplir para cada $k \in C_i^s$ la siguiente igualdad:

$$f_k(L_j^k, 0) = \sum_{l\in C_i^e} \alpha_{lk}^i f_l(R_j^l, L_l). \tag{24}$$

Para que la igualdad anterior se cumpla, es necesario que la información en la frontera se imponga sobre la información en el interior de la carretera y una condición para que esto sea así, fue establecida en (12) y (13). En términos del sistema de carreteras esto sería, para todo $k \in C_i^s$:

$$L_j^k \in [0, \min\{{}_k\overline{Q}_0^j, q_{opt}^k(1-\beta_k)\}], \tag{25}$$

y para todo $l \in C_i^e$:

$$R_j^l \in [\max\{{}_l\overline{Q}_n^j, q_{opt}^l(1+\beta_l)\}, q_{max}^l]. \tag{26}$$

Dado que el flujo queda determinado de manera unívoca en los intervalos descritos en (25) y (26) la ecuación (24) se puede reformular en términos de flujos. Defínanse ${}_k\eta_{max}^j = f_k(\min\{{}_k\overline{Q}_0^j, q_{opt}^k(1-\beta_k)\}, 0)$, ${}_l\xi_{max}^j = f_l(\max\{{}_l\overline{Q}_n^j, q_{opt}^l(1+\beta_l)\}, L_l)$, $\eta_j^k = f_k(L_j^k, 0)$ y $\xi_j^l = f_l(R_j^k, L_l)$, en términos de estas variables la ecuación (24) quedaría expresada de la siguiente forma:

$$\eta_j^k = \sum_{l \in C_i^e} \alpha_{lk}^i \xi_j^l, \tag{27}$$

donde $\eta_j^k \in [0, {}_k\eta_{max}^j]$ y $\xi_j^l \in [0, {}_l\xi_{max}^j]$. Para poder encontrar tales flujos, en [5] se teoriza que los flujos deben maximizarse en las intersecciones bajo una cierta regla, esta regla podría ser directamente la suma de los flujos, o como se plantea en dicho artículo, una maximización a través de una función cóncava. En este trabajo se usa la suma de los flujos, la cual da lugar a un problema de optimización lineal. Formalmente se define a continuación, el problema que se quiere resolver:

Definición 3. *Sean $i \in V$ con $grad(i) > 2$ y el intervalo de tiempo $[t_j, t_{j+1}]$. Considérese encontrar $(\eta_j^k)_{k \in C_i^s}$ y $(\xi_j^l)_{l \in C_i^e}$ de manera que se maximice la siguiente expresión:*

$$\sum_{k \in C_i^s} \eta_j^k + \sum_{l \in C_i^e} \xi_j^l,$$

donde para todo $k \in C_i^s$ se cumple la ecuación (27)*.*

El problema dado en la Definición 3 se puede escribir como un problema de programación lineal: Sean $C_i^s = \{k_1, k_2, \cdots, k_u\}$, $C_i^e = \{l_1, l_2, \cdots, l_v\}$, $\xi_i = (\xi_j^{l_1}, \xi_j^{l_2}, \cdots, \xi_j^{l_v})^T$, $A \in \mathbb{R}^{u \times v}$ donde la entrada (c, d) es $\alpha_{l_d k_c}$, $w = ({}_{k_1}\eta_{max}^i, {}_{k_2}\eta_{max}^i, \cdots, {}_{k_u}\eta_{max}^i, {}_{l_1}\xi_{max}^i, {}_{l_2}\xi_{max}^i, \cdots, {}_{l_v}\xi_{max}^i, 0, \cdots, 0)^T$, $w \in \mathbb{R}^{u+2v}$. Con todo lo anterior se define el problema siguiente:

$$Maximizar: \ \xi_i$$

$$Sujeto\ a: \begin{pmatrix} A \\ I^{v \times v} \\ -I^{v \times v} \end{pmatrix} \xi_i \leq w$$

Como se puede apreciar, el problema anterior es un programa lineal, el cual se puede resolver usando la vasta teoría sobre optimización lineal.

5. Experimentos numéricos

Esta sección tiene como objetivo validar la implementación de un código elaborado en *python* y los resultados teóricos expuestos en (23) y (19). Para poder hacer tal experimento se ha usado un ejemplo del cual se conoce la solución exacta.

Ejemplo 1. *Considere el problema* (5) *donde* $a = -10$, $b = 120$ *para* $t \in [0,1]$ *y* $x \in [0,1]$, *sujeto a las condiciones* $q(x,0) = q_0$ *para* $x < \frac{1}{2}$ *y* $q(x,0) = q_1$ *para* $x > \frac{1}{2}$, *donde* q_0 *y* q_1 *son constantes.*

La solución al Ejemplo (1) viene dada por la siguiente expresión:

$$q(x,t) = \begin{cases} \dfrac{q_0 q_{max}}{(q-q_0)e^{at}+q_0} & para\ x(t) < \dfrac{1}{2} \\ \dfrac{q_1 q_{max}}{(q-q_1)e^{at}+q_1} & para\ x(t) > \dfrac{1}{2} \end{cases},$$

donde $x(t) = \dfrac{\left(q_{max}e^{\frac{at}{2}} - 2q_0\sinh(\frac{at}{2})\right)\left(q_{max}e^{\frac{at}{2}} - 2q_1\sinh(\frac{at}{2})\right) - bq_{max}^2}{aq_{max}^2}$.

Denótese por Q la solución aproximada usando el flujo numérico de Godunov descrito en el marco computacional para el problema descrito en el Ejemplo 1. Usando esta aproximación es posible medir el error de la siguiente forma:

$$Error = ||q - Q||_{1,h} = \Delta x \sum_{i=0}^{n} |q_i - Q_i^m|, \tag{28}$$

donde:

$$q_i = \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} q(x,T)dx.$$

Usando esta definición y el código implementado, se obtuvo la Tabla 1 donde se hicieron tres experimentos, en el primero se fijo la densidad máxima en 120, en el segundo se fijo en 140 y en el tercero 100. La Tabla 1 posee tres secciones; "Región de no estabilidad", "Región de estabilidad" y "Condición suficiente", para la primer sección se escogieron pares $(\Delta x, \Delta t)$ de manera que no se cumpliera la condición (19), en la segunda estos pares verifican la condición (19) y en la última se verifica el resultado de la Proposición 4. Para cada sección y experimento se utilizó una relación de la forma $\Delta x = A\ln(1 + B\Delta x)$ en dirección con el resultado encontrado en la Proposición 4.

Como se puede apreciar en la Tabla 1, se validan lo resultados teóricos encontrados con la condición (19) y la Proposición 4. Particularmente se puede notar que en la región de no estabilidad, a pesar de que los valores Δt y Δx se hacen pequeños, aun así, no se logra alcanzar la convergencia del método.

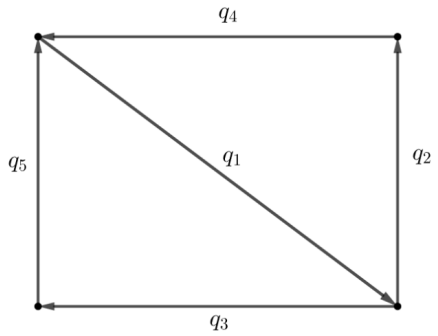
En [7] se puede encontrar (capítulo de stabilidad no lineal) que la tasa de convergencia del método de volúmenes finitos usando el flujo de Godunov, es coherente con la tasa encontrada en los experimentos.

Tabla de convergencia											
Δt	Δx	$Error$	$Tasa$	Δt	Δx	$Error$	$Tasa$	Δt	Δx	$Error$	$Tasa$
Experimento 1				Experimento 2				Experimento 3			
Región de no estabilidad											
4.25e-03	1.25e-01	NaN	NaN	2.52e-03	1.25e-01	NaN	NaN	4.245e-03	1.25e-01	NaN	NaN
2.11e-03	6.25e-02	NaN	NaN	1.26e-03	6.25e-02	NaN	NaN	2.108e-03	6.25e-02	NaN	NaN
1.05e-03	3.12e-02	NaN	NaN	6.27e-04	3.12e-02	NaN	NaN	1.047e-03	3.125e-02	NaN	NaN
5.23e-04	1.56e-02	NaN	NaN	3.13e-04	1.56e-02	NaN	NaN	5.233e-04	1.562e-02	NaN	NaN
2.6e-04	7.81e-03	NaN		1.57e-04	7.81e-03	NaN		2.599e-04	7.812e-03	NaN	
Región de estabilidad											
1.15e-03	1.25e-01	1.04e-05	1.0	1.05e-03	1.25e-01	1.85e-05	1.0	1.145e-03	1.25e-01	1.036e-05	1.0
5.69e-04	6.25e-02	5.17e-06	1.0	5.23e-04	6.25e-02	9.21e-06	1.0	5.687e-04	6.25e-02	5.169e-06	1.0
2.84e-04	3.12e-02	2.58e-06	1.0	2.6e-04	3.12e-02	4.60e-06	1.0	2.844e-04	3.125e-02	2.582e-06	1.0
1.42e-04	1.56e-02	1.29e-06	1.0	1.3e-04	1.56e-02	2.3e-06	1.0	1.422e-04	1.562e-02	1.290e-06	1.0
7.11e-05	7.81e-03	6.45e-07		6.5e-05	7.81e-03	1.15e-06		7.109e-05	7.812e-03	6.450e-07	
Condición suficiente											
1.05e-03	1.25e-01	9.49e-06	1.0	8.99e-04	1.25e-01	1.58e-05	1.0	1.047e-03	1.25e-01	9.491e-06	1.0
5.23e-04	6.25e-02	4.74e-06	1.0	4.46e-04	6.25e-02	7.89e-06	1.0	5.233e-04	6.25e-02	4.737e-06	1.0
2.6e-04	3.12e-02	2.37e-06	1.0	2.23e-04	3.12e-02	3.94e-06	1.0	2.599e-04	3.125e-02	2.367e-06	1.0
1.3e-04	1.56e-02	1.18e-06	1.0	1.12e-04	1.56e-02	1.97e-06	1.0	1.299e-04	1.562e-02	1.183e-06	1.0
6.5e-05	7.81e-03	5.91e-07		5.58e-05	7.81e-03	9.85e-07		6.496e-05	7.812e-03	5.913e-07	

Tabla 1: Información sobre la convergencia del método de volúmenes finitos.

Con ayuda del marco computacional se construyó un código en python para encontrar la solución al modelo descrito en la Definición 1. Los dos siguientes experimentos validan el código implementado.

Ejemplo 2. *Considérese una red tráfico vehicular con 5 carreteras, como se muestra a continuación:*



Sujeto a las siguientes condiciones:

1. *Condiciones iniciales:* $q_2(x,0) = q_3(x,0) = q_4(x,0) = q_5(x,0) = 0$ *y* $q_1(x,0) = 110e^{-15(2x-1)^2}$.
2. *Condiciones de frontera: El flujo de la carretera correspondiente a la densidad* q_1 *se divide por igual en las carreteras correspondientes a las densidades* q_2 *y* q_3.
3. *Con referencia a la Definición 1, se tiene que* $a_l = -10$, $b_l = 70$ *y* $q^l_{max} = 120$ *para todas las carreteras.*

4. *La longitud de cada carretera es de un kilómetro y la dinámica se estudia durante seis minutos.*

Al utilizar el código implementado en este trabajo, se obtuvo la evolución de la densidad en el tiempo para este grafo de carreteras, esto se muestra en la Figura 2.

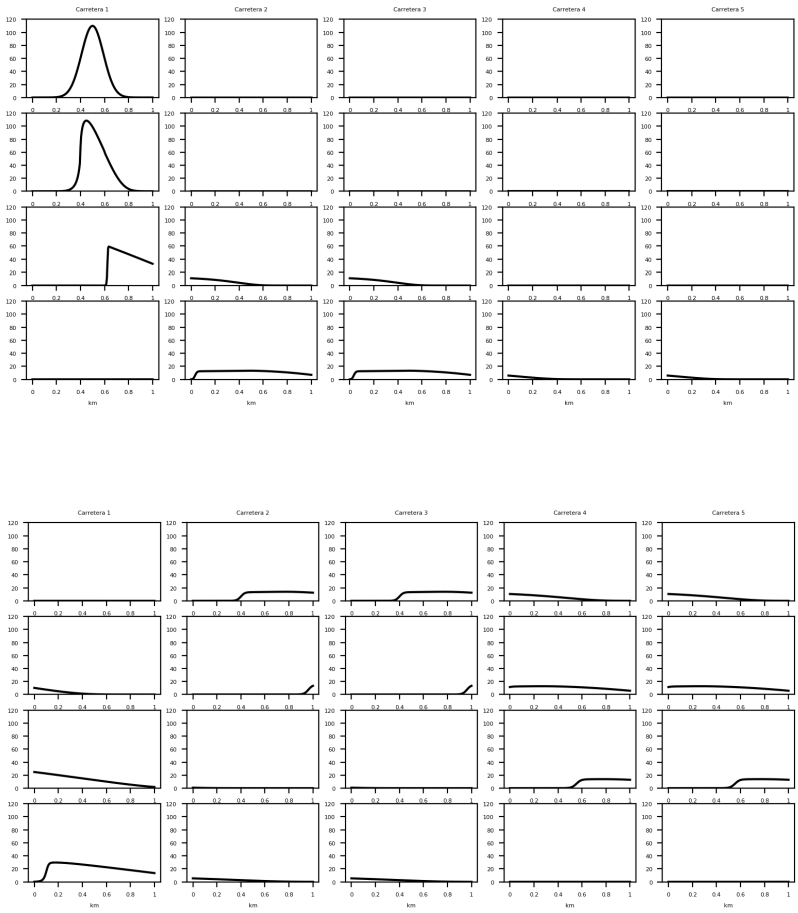


Figura 2: En las imágenes se muestra la evolución de la densidad en cada una de las carreteras en los tiempos t =0, 4, 43, 86, 108, 144, 180 y 216 (medidos en segundos) de arriba a abajo respectivamente.

Al igual que en el Ejemplo 1, aquí se desea medir el error de la solución encontrada. Una de las dificultades para medir el error en este ejemplo, se da debido a que no es sencillo conseguir una solución analítica del problema (Ejemplo 2). Una de las formas en las que se puede medir la solución en este tipo de casos, consiste en hacer una partición bastante fina en el tiempo y considerar a esta aproximación numérica como si fuese la solución, al hacer esta suposición se espera que con particiones relativamente más gruesas se de un comportamiento similar al contrastar estás soluciones con la solución verdadera. Además este tipo de experimentos siempre debe ir acompañado de observar un comportamiento apropiado en las

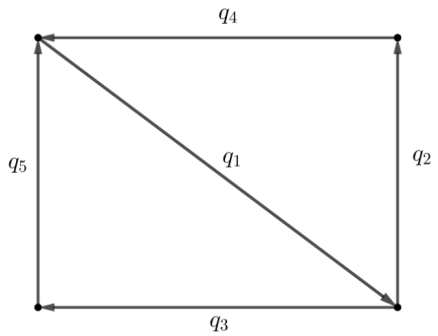
soluciones numéricas; para este ejemplo se puede afirmar que las soluciones tienen un comportamiento esperado, basándose en lo observado en la Figura 2.

En la Tabla 2 se puede observar el error definido en (28) para cada una de las carreteras, además se calcularon las tasas, producto de las diferentes aproximaciones. Se puede observar que en cada una de las carreteras el error tiene un comportamiento decreciente a medida que Δt se hace más fino y el valor aproximado de la tasa es coherente con los resultados encontrados en el Ejemplo 1.

Tabla de convergencia										
Δt	*Error* 4	*Error* 2	*Error* 3	*Error* 1	*Error* 5	*Tasa* 4	*Tasa* 2	*tasa* 3	*Tasa* 1	*tasa* 5
5.e-05	2.5e-01	1.0e-01	1.0e-01	9.7e-02	2.5e-01	1.07	1.06	1.06	1.04	1.07
2.5e-05	1.2e-01	4.9e-02	4.9e-02	4.7e-02	1.2e-01	1.04	1.04	1.04	1.03	1.04
1.3e-05	5.8e-02	2.4e-02	2.4e-02	2.3e-02	5.8e-02	1.04	1.03	1.03	1.03	1.04
6.3e-06	2.8e-02	1.2e-02	1.2e-02	1.1e-02	2.8e-02	1.06	1.05	1.05	1.05	1.06
3.1e-06	1.4e-02	5.6e-03	5.6e-03	5.5e-03	1.4e-02	1.11	1.1	1.1	1.1	1.11

Tabla 2: Información

Ejemplo 3. *Considérese una red tráfico vehicular con 5 carreteras, como se muestra a continuación:*



Sujeto a las siguientes condiciones:

1. *Condiciones iniciales:* $q_2(x,0) = q_3(x,0) = q_4(x,0) = q_5(x,0) = 0$ *y* $q_1(x,0) = 110e^{-15(2x-1)^2}$.
2. *Condiciones de frontera: El flujo de la carretera correspondiente a la densidad* q_1 *se divide por las carreteras con densidades* q_2 *y* q_3 *en los porcentajes de 35 y 65 por ciento respectivamente.*
3. *Con referencia a la Definición 1, se tiene que:*

$$\begin{aligned}(a_1, a_2, a_3, a_4, a_5) &= (-20, -30, 10, 5, -10),\\ (b_1, b_2, b_3, b_4, b_5) &= (80, 100, 90, 100, 70),\\ (q^1_{max}, q^2_{max}, q^3_{max}, q^4_{max}, q^5_{max}) &= (120, 100, 130, 80, 140).\end{aligned}$$

4. *La longitud de cada carretera es de un kilómetro y la dinámica se estudia durante doce minutos.*

Tabla de convergencia										
Δt	*Error* 4	*Error* 2	*Error* 3	*Error* 1	*Error* 5	*Tasa* 4	*Tasa* 2	*tasa* 3	*Tasa* 1	*tasa* 5
5.e-05	5.1e-03	9.7e-02	5.0e-02	2.9e-01	1.7e-02	1.1	0.97	0.98	1.07	1.09
2.5e-05	2.4e-03	4.9e-02	2.6e-02	1.4e-01	8.0e-03	1.04	1.01	1.01	1.04	1.03
1.3e-05	1.1e-03	2.4e-02	1.3e-02	6.7e-02	3.9e-03	1.07	1.04	1.04	1.06	1.06
6.3e-06	5.5e-04	1.2e-02	6.2e-03	3.2e-02	1.9e-03	1.11	1.1	1.1	1.11	1.11
3.1e-06	2.5e-04	5.6e-03	2.9e-03	1.5e-02	8.7e-04	1.23	1.23	1.23	1.23	1.23

Tabla 3: Información

A diferencia del Ejemplo 2, en este experimento se impusieron condiciones asimétricas. El comportamiento decreciente del error se sigue presentando aun en estas condiciones como se muestra en la Tabla 3.

6. Conclusiones

En este trabajo se logró desarrollar un marco computacional para un modelo del tráfico vehicular en una red de carreteras desde la visión de los modelos macroscópicos, considerando un flujo dependiente linealmente de la variable espacial. Para la elaboración del marco computacional se utilizó el método de volúmenes finitos. Particularmente se calculó el flujo numérico de Godunov para el tipo de ecuación diferencial parcial presente en el modelo.

Se encontraron dos condiciones, una suficiente y otra necesaria, para garantizar la estabilidad del método de volúmenes finito y se validaron tales resultados teóricos por medio de algunos experimentos numéricos.

Se encontraron las condiciones sobre el flujo en las intersecciones similares a las expuestas por Holden y Risebro, ahora considerando el flujo dependiente linealmente de la variable espacial.

Se validó el código implementado en *python* por medio de dos experimentos sobre una red de carreteras en un circuito cerrado.

Referencias

[1] e. a. Greenshields, B. D., A study of traffic capacity, in: Highway research board proceedings, volume 1935, National Research Council (USA), Highway Research Board, 1935.

[2] G. B. Lighthill, M. J. y Whitham, On kinematic waves i. flood movement in long rivers, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 229 (1955) 281–316.

[3] G. B. Lighthill, M. J. y Whitham, On kinematic waves ii. a theory of traffic flow on long crowded roads, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 229 (1955) 317–345.

[4] P. I. Richards, Shock waves on the highway, Operations research 4 (1956) 42–51.

[5] N. H. Holden, H. y Risebro, A mathematical model of traffic flow on a network of unidirectional roads, SIAM Journal on Mathematical Analysis 26 (1995) 999–1017.

[6] G. M. y. P. B. Coclite, G. M., Traffic flow on a road network, SIAM journal on mathematical analysis 36 (2005) 1862–1886.

[7] R. J. LeVeque, Randall J y Leveque, Numerical methods for conservation laws, volume 132, Springer, 1992.

MODELOS OCULTOS DE MARKOV PARA LA DETECCIÓN DE FALLAS

DAVID ARMANDO ORDOÑEZ TABORA

Resumen. Los modelos ocultos de Markov (MOM) son procesos doblemente estocásticos donde se tiene un conjunto de observaciones, también llamadas señales, que provienen de un conjunto de estados ocultos. En ese sentido, se tiene una distribución de probabilidad entre estados, A, una distribución de probabilidad de las observaciones, B, y una distribución de estado inicial, π, denotando el modelo como $\lambda(A, B, \pi)$. Estos modelos se han aplicado para la detección de fallas en sistemas dinámicos, ya que permiten modelar la evolución temporal de un sistema y así poder detectar anomalías que sean indicativas de un fallo. Además, son herramientas que ayudan a filtrar datos ruidosos, e inclusive, trabajar con datos incompletos. Por ello, en este trabajo se implementarán los MOM para problemas de detección de fallas. Analizando las características y tareas potenciales de estos frente al fenómeno estudiado. Se expone una metodología para poder lograr la implementación de los MOMs, considerando estrategias para determinar aspectos necesarios en torno a ellos como ser, la cantidad de estados ocultos, la topología entre los estados ocultos y la distribución inicial.

Abstract. Hidden Markov Models (HMM) are doubly stochastic processes where there is a set of observations, also called signals, that come from a set of hidden states. In this sense, we have a probability distribution between states, A, a probability distribution of observations, B, and an initial state distribution, π, denoting the model as $\lambda(A, B, \pi)$. These models have been applied to detect failures in dynamic systems, since they allow modeling the temporal evolution of a system and thus be able to detect anomalies that are indicative of a failure. In addition, they are a tool that can filter out noisy or incomplete data in order to identify the underlying signal in the data. Therefore, in this work the HMM will be implemented for fault detection problems. Analyzing the characteristics and potential tasks of these in front of the studied phenomenon. A methodology is exposed to be able to achieve the implementation of the HMMs, considering strategies to determine necessary aspects around them, such as the number of hidden states, the topology between the hidden states and the initial distribution.

1. Introducción

En ingeniería matemática, los problemas de detección de fallas conllevan una serie de técnicas y modelos con el fin de poder determinar a tiempo la falla para realizar acciones correctivas que minimicen el impacto de estas. Una falla se define como el cambio de comportamiento de algunos componentes de un sistema, donde ya no cumplen con la labor correcta en el proceso [4]. En ese sentido, resulta necesario emplear técnicas y modelos adecuados al proceso para lograr una detección

Date: 19 de agosto de 2023.
Palabras claves. Detección de fallas, Modelos ocultos de Markov, Aprendizaje automático.

temprana de la falla.

Entre los modelos que se han implementado para detectar anomalías en un proceso se encuentran los modelos ocultos de Markov (MOM) [6, 8, 12]. Los cuales se definen como un proceso bivariado de tiempo discreto $\{X_k, Y_k\}_{k\geq 0}$, donde $\{X_k\}$ es una cadena de Markov, $\{Y_k\}$ es una sucesión de variables aleatorias independientes, tales que la distribución condicional de Y_k solamente depende de X_k. El espacio de estados de la cadena de Markov, $\{X_k\}$, se denota por X y el conjunto de valores que toma ,$\{Y_k\}$, por Y [2].

Los modelos ocultos de Markov, dadas sus características, son métodos que se pueden usar para representar problemas con propiedades temporales inherentes, por ejemplo, problemas donde se tienen eventos secuenciales. Particularmente, los procesos de maquinarias tienen esta característica, resultando así en la implementación de MOMs para lograr determinar una posible falla, e inclusive, clasificar de que estado proviene la falla. Esto indica potencialidades de los MOMs y como se plantea en [11], se debe explorar la aplicación de ellos en la industria.

En el presente trabajo, se busca exponer una metodología para implementar los MOMs a problemas de detección de fallas, explorando los aspectos teóricos de ellos y haciendo variaciones para determinar condiciones que permitan mejorar la eficiencia de los modelos frente al problema propuesto. Lo anterior se logra mediante un análisis profundo a la literatura y trabajos donde se han aplicado los MOMs, con ello se determinan estrategias para obtener una cantidad adecuada de estados ocultos, optimizar los parámetros del modelo e inclusive, hacer variaciones al modelo, como debilitar la propiedad de Markov a una propiedad semiMarkoviana [13] logrando así que se considere una evolución temporal del modelo.

2. Justificación

La detección de fallas se refiere a la identificación de anomalías de un dispositivo, maquinaria o proceso. El cual corresponde a un aspecto vital para evitar pérdidas económicas, paros en procesos o deterioro de maquinarias que conllevan a costos grandes en reparación o, inclusive, a la pérdida total de la máquina. Aunado a ello, es parte de uno de los aspectos a considerar en cuanto a la seguridad industrial [9]. Por lo que, resulta relevante abordar técnicas que permitan detectar fallas a tiempo para iniciar acciones que busquen corregir la misma.

Dentro de las técnicas para la detección y diagnóstico de fallas que se suelen emplear se destacan aquellas que consideran datos históricos del proceso y así establecen un modelo que permita determinar cuando se puede presentar una falla [4]. Estas técnicas son propias del aprendizaje automático, como redes neuronales o MOMs. Particularmente, los MOMs, por sus características de representación de secuencias y la exploración de estados ocultos resultan beneficiosos para la identificación de anomalías.

La implementación de estos modelos para el problema planteado es parte de una necesidad que posee el sector industrial e instituciones gubernamentales que

conlleven procesos, donde una falla en ellos genera no solo pérdidas cuantiosas sino insatisfacción e inoperancia. Por tal razón, el presente trabajo muestra potencial práctico de interés a los sectores del país descritos anteriormente.

Finalmente, considerando las líneas de investigación de la *Universidad Nacional Autonóma de Honduras* (UNAH) el estudio corresponde a los ejes de investigación *Desarrollo económico y social, Población y condiciones de vida* y *Ambiente, biodiversidad y desarrollo*, esto en función de donde se implemente. Además, dentro de las líneas de investigación de la *Maestría en Matemática con Orientación en Ingeniería Matemática*, corresponde a las líneas de: autómatas celulares, por ser los autómatas probabilísticos una generalización de los MOMs [3]; modelación matemática, ya que se relaciona con la teoría de control y; simulación computacional, ya que se analiza el rendimiento, precisión y eficacia del MOM empleado mediante la implementación en python.

3. Antecedentes

Los MOMs son modelos que surgieron desde hace 60 años, a raíz de buscar caracterizar procesos estocásticos para los cuales no se contaban con suficientes observaciones. En ese sentido, con las observaciones que se tenían se consideraba que estas se regían por otro proceso que estaba oculto, logrando así caracterizar el proceso oculto y otro que brindaba las observaciones solamente con lo que se podía observar [8]. Los trabajos publicados por *Leonard E. Baum* en conjunto con otros investigadores, entre los años de 1960 a 1970, fueron los primeros donde se introducen los MOMs. A partir de allí, muchos otros investigadores han hecho uso de estos modelos aplicados a diversas áreas.

Dentro de los trabajos donde se han empleado los MOMs, *Mor* et al. en [8], muestran variantes de estos, así como las áreas en que se han aplicado. La flexibilidad de los MOMs y el desarrollo de nuevas estrategias, han permitido variaciones en función de la necesidad o adecuación para cada problema estudiado. Entre las variaciones que se establecen se encuentran, modelos ocultos de Markov de alto orden ($AO\text{-}MOM$), modelos ocultos semi-Markovianos ($MOSM$), modelo factorial oculto de Markov ($MFOM$), modelo oculto de Markov en capas ($MOMC$), modelo autoregresivo oculto de Markov ($MAROM$), modelo oculto de Markov no estacionario ($MOM\text{-}NE$), modelo oculto de Markov jerárquico ($MOMJ$), entre otros.

Por otro lado, los MOMs se han aplicado en áreas como biología, finanzas, ingeniería, salud, entre otras. Particularmente en el área de ingeniería se han usado para predecir el deterioro o fallas de maquinarias [3, 6, 12]. Por consiguiente, a continuación se explora el uso de estos para la detección de fallas.

La investigación realizada por *Kouadri et al.* en [6], plantea que la alta aleatoriedad del entorno operativo del convertidor de los sistemas de conversión de energía eólica, dificulta la detección y diagnóstico de fallas. Para ello hacen uso de un MOM basado en Análisis de Componentes Principales (PCA), con el fin que el PCA extraiga y seleccione de manera eficiente las características y estados para el MOM. En ese sentido, el MOM lo que hace es clasificar diferentes fallas que pueden ocurrir en los convertidores de potencia. El proceso que plantean los autores muestra la

necesidad de analizar la estructura del convertidor de potencia y las conexiones a él para determinar la topología del modelo. Luego aplican *PCA* para determinar las características significativas y con ellas alimentan al MOM para que clasifique las fallas. Los resultados mostraron que el modelo presentado elimina la dificultad de detección de fallas ya que se disminuye el ruido a las señales que la aleatoriedad del entorno ocasiona, resultando mejor que la Máquina de Soporte Vectorial (MSV) comúnmente empleado para estos problemas.

Por otro lado, el trabajo presentado por *Wu et al.* en [12], considera una variación de un MOM para estimar el deterioro de una máquina y predecir la vida útil restante basado en múltiples indicadores claves de rendimiento. El estudio, a la luz de la literatura consultada, propone el MOM basado en clusters de dos fases para aprovechar el agrupamiento, considerando colocar en un solo marco los enfoques de MOM multicapa y MOM basado en patrones. El método empleado para los clusters consiste en usar la técnica de agrupamiento espacial y el algoritmo K-means. El primero permite inicializar la secuencia de observaciones de símbolos y genera una cantidad inicial de estados de deterioro, mientras que el segundo toma la cantidad inicial de estados anteriores y hace una comparación para ajustar la cantidad de estados de deterioro. Con ello entrenan el MOM y lo aplican para cada estado, similar al trabajo de [3] donde consideran un modelo para cada clase. La razón es para tener una forma de reconocer de donde viene el problema de deterioro. Usan el algoritmo del *mejor símbolo* y lo aplican a un *algoritmo extendido de Viterbi*, todo ello para reconocer el estado actual de la máquina, comparándolo con la probabilidad de secuencia dada por uno de los MOM de cada estado. Finalmente, el modelo genera el estado actual de deterioro y la raíz del mismo, y aplicando log-verosimilitud determinan el peso de los parámetros y con ello estiman la vida útil restante de la maquinaria. Es así como logran considerar una variante de MOM eficiente, que es mejor que casi todos los otro métodos con los que lo comparan, excepto por uno, el *deep-lstm*.

El trabajo de *Tai, Ching y Chan* en [11], hacen uso de un MOM para determinar que máquina presenta fallo en el llenado de una botella. En este trabajo se tienen varias máquinas de llenado, y solo se observan si las botellas fueron llenadas dentro del rango establecido o no, lo que corresponde a las señales (observaciones) del modelo. El número de estados corresponden a cada máquina y buscan predecir que máquina fallará. Luego de las experimentaciones realizadas y los resultados obtenidos concluyeron que el MOM implementado es efectivo para detectar la máquina que está fallando. Asimismo, sugieren que en caso la cantidad de máquinas sea muy grande se considere establecer una partición en subproblemas o hacer uso de un AO-MOM.

Por último, el trabajo de *Arapaia et. al.* en [1], donde aplican un MOM para detectar fallas en maquinarias de fluidos, expone potencialidades de estos modelos, ya que, en el estudio los datos a priori que poseen no tienen la información adecuada y solo entrenan el modelo con datos adquiridos durante el funcionamiento normal de la máquina. Además, considerando las medidas que se tienen de la máquina de fluido toman un MOM para cada caso, identificando las características (estados ocultos) mediante un análisis de componentes principales. Concluyen con la evaluación y validación del modelo las cuales fueron satisfactorias y proponen que se

apliquen estos modelos a otros tipos de maquinarias.

En resumen, luego de la exploración de trabajos en los cuales han implementado los MOMs para detección de fallas, se muestra la necesidad de hacer uso de técnicas, como el análisis de componentes principales y estrategias de agrupamiento, para poder extraer las características a considerar en el modelo y los estados ocultos. Además, una forma de decidir cuando hacer uso de los MOMs es analizar la información que se tiene, donde, a pesar que esté incompleta o no sea adecuada, estos pueden trabajar con ellas, ya que logran estimar la información faltante. Por otro lado, todos los trabajos proponen considerar diferentes tipos de maquinarias o contextos para implementar estos modelos y sus variantes, y así descubrir bajo que aspectos los MOMs se consideran mejores frente a otros modelos de detección de fallas.

4. Preliminares

Primeramente, para la implementación de los MOMs, es necesario conocer los aspectos teóricos en torno a ellos para lograr considerar las variaciones en función de las características del problema a analizar. En la siguiente sección se expone la teoría y algoritmos a usar para estos modelos.

4.1. Definición de los MOMs. Un MOM consta de ciertos elementos particulares que nos remiten a su definición. Estos elementos se basan en los dos procesos estocásticos que involucran los MOMs, uno que no brinda datos observables (oculto) y otro que genera las observaciones. Considerando esto, a continuación se definen los 5 elementos generales [5]:

1. Un conjunto de N estados, dados por $S = \{S_1, ..., S_N\}$.
2. Un conjunto de M símbolos, que brindan las observaciones por estado, definido como $V = \{v_1, ..., v_M\}$. En caso que las observaciones sean continuas entonces M es infinito.
3. La distribución de probabilidad de transición entre estados, dada por $A = \{a_{ij}\}$ donde a_{ij} es la probabilidad de estar en el tiempo t en el estado i y pasar al estado j en el tiempo $t+1$. En ese sentido, la estructura de la matriz estocástica definirá las conexiones del modelo. Formalmente se define como

$$a_{ij} = P\{q_{t+1} = j | q_t = i\},$$

   donde q_t denota el estado actual.
4. La distribución de probabilidad de observaciones de cada estado dada por $B = \{b_j(k)\}$, donde $b_j(k)$ es la probabilidad que el símbolo v_k sea emitido por el estado S_j, más precisamente

$$b_j(k) = P\{o_t = v_k | q_t = j\},\ 1 \leq j \leq N,\ 1 \leq k \leq M,$$

   donde v_k es el $k-$ésimo símbolo de observación del conjunto M y o_t el vector actual de parámetros. En caso que las observaciones sean continuas, se debe hacer uso de una función de densidad de probabilidad continua. En ese caso se debe especificar los parámetros de la función de densidad de probabilidad. Generalmente la densidad de probabilidad se aproxima por

una suma ponderada de M distribuciones Gaussianas N, como sigue

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} N(\mu_{jm}, \Sigma_{jm}, o_t)$$

donde c_{jm} son los coeficientes de ponderación, μ_{jm} vectores de medias y Σ_{jm} la matriz de covarianza. Además c_{jm} debe cumplir que $c_{jm} \geq 0$, $1 \leq j \leq N$, $1 \leq m \leq M$ y $\sum_{m=1}^{M} c_{jm} = 1$, $1 \leq j \leq N$.

5. La distribución de estado inicial del MOM, denotada por, $\pi = \{\pi_i\}$ donde π_i es la probabilidad que el modelo esté en el estado S_i en el tiempo inicial, con $\pi_i = P\{q_t = i\}$ y $1 \leq i \leq N$.

Con lo anterior, se define un MOM como un proceso doblemente estocástico $\lambda = (A, B, \pi)$, en el caso sea discreto, y $\lambda = (A, c_{jm}, \mu_{jm}, \Sigma_{jm}, \pi)$ en caso sea continuo.

Note que para representar un MOM continuo, la distribución de probabilidad de observaciones, B, se representa considerando los parámetros de la aproximación de la suma Gaussiana. En ese sentido, al momento de trabajar con estos modelos es importante caracterizar el tipo del mismo. Entre la categorización en función del tiempo se podría tener un MOM discreto, continuo o mixto. Pero también se puede caracterizar considerando otros aspectos como el orden de la cadena de Markov oculta, o inclusive debilitar la condición de Markov y considerar un modelo oculto semimarkoviano MOSM [13]. Sin embargo, todos implican los 5 elementos anteriormente descritos para su definición.

Por otro lado, de acuerdo a lo planteado se tiene que el proceso estocástico oculto, dado por el conjunto de estados y la transición en éstos, cumple la propiedad markoviana, o en el peor de los casos, la propiedad semimarkoviana. La propiedad markoviana se resume en que la probabilidad de estar en el estado j partiendo del estado i solo depende del tiempo anterior a este, mientras que la propiedad semimarkoviana establece que la probabilidad de ir del estado i al estado j en el tiempo $t + 1$ depende de la cantidad de tiempo que ha transcurrido desde que entró al estado actual t.

Con respecto al proceso estocástico observable, el cual brinda las observaciones, se tiene que el conjunto de símbolos que se generan solo dependen del conjunto de estados S, es decir, del estado que brinda esa observación. Por lo que el tipo de MOM en cuanto al tiempo lo determina el proceso observable. En ese sentido, el modelo oculto se considera un modelo de estados finito. Sin embargo, esta no es una condición estricta, ya que en un sentido más general, como plantea *Bickel et al.* en [2], se puede tener que la probabilidad de observaciones dependa no solamente de los estados que las generan en el tiempo actual, sino también de las observaciones pasadas, dando paso a lo que se conoce como un modelo oculto de conmutación de Markov. Estos aspectos ya tienen que ver con la arquitectura del modelo, o como mejor se conoce, la topología del modelo.

4.2. Algoritmos de los *HMM*. En el desarrollo de los MOMs se presentaron tres problemas fundamentales que se deben considerar para su implementación en el mundo real. Para ello se formularon algoritmos que permitieron solventar los problemas. Hoy en día con los avances en tecnología y ciencia se han mejorado estos

algoritmos y creado otros, sin embargo, los primeros, siguen siendo fundamentales y la base para los nuevos algoritmos o técnicas que se han presentado para solventar los problemas.

A continuación se hace una breve descripción de los tres problemas fundamentales.

4.2.1. Problema de evaluación. El problema de evaluación busca responder lo siguiente, ¿Cuál es la probabilidad de que dada una observación $O = o_1, ..., o_T$, esa observación sea generada por el modelo $P\{O|\lambda\}$ donde λ es el MOM?

La respuesta a la interrogante puede ser encontrada con cálculos simples, pero el problema de ello es la gran cantidad de operaciones que implicarían, de acuerdo a [5] la cantidad de operaciones es de orden N^T, donde N es la cantidad de estados y T el tiempo transcurrido. La forma para reducir el número de cálculos se encuentra en el uso del algoritmo *forward-backward*.

El algoritmo, extraído de [10] y [5], plantea lo siguiente:
Considera una variable hacia adelante $\alpha_t(i) = P\{o_1, ..., o_t, q_t = i|\lambda\}$, esto es, la probabilidad de tener una sucesión de observaciones parciales hasta el tiempo t y el estado i en el tiempo t, dado el modelo λ. Luego se obtiene la siguiente fórmula recursiva

$$\alpha_{t+1}(j) = b_j(o_{t+1}) \sum_{i=1}^{N} \alpha_t(i) o_{ij},\ 1 \leq j \leq N,\ 1 \leq t \leq T-1$$

con $\alpha_1(j) = \pi_j b_j(o_1),\ 1 \leq j \leq N$.

Luego, $\alpha_T(i)$ para $1 \leq i \leq N$ se calcula con la fórmula recursiva anterior.

Finalmente la probabilidad requerida estará dada por $P\{O|\lambda\} = \sum_{i=1}^{N} \alpha^T(i)$. Hasta este momento se tiene el procedimiento *hacia adelante*.

Para el procedimiento hacia atrás (*backward*) se trabaja de forma similar, solo que en este caso se tendría una variable hacia atrás dada por

$$\beta_t(i) = P\{o_{t+1}, o_{t+2}, ..., o_T|q_t = i, \lambda\}.$$

Cada $\beta_t(i)$ se calcula aplicando la fórmula recursiva siguiente

$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j) a_{ij} b_j(o_{t+1}),\ 1 \leq i \leq N,\ 1 \leq t \leq T-1$$

donde $\beta_T(i) = 1$, $1 \leq i \leq N$.

Hasta este momento se tiene que $\alpha_t(i)$ determina la probabilidad de la secuencia parcial de $o_1, ..., o_t$ y el estado i en el tiempo t, mientras que $\beta_t(i)$ determina la probabilidad de la secuencia parcial $o_{t+1}, ..., o_T$, dado que está en el estado i al tiempo t. Por lo que establece una partición de la trayectoria para obtener la observación $O = o_1, ..., o_T$. Entonces usando ambas variables se obtiene que

$$P\{O|\lambda\} = \sum_{i=1}^{N} P\{O, q_t = i|\lambda\} = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i).$$

Al algoritmo descrito anteriormente, también se le conoce como algoritmo de *Baum-Welch* para los MOMs [13], la idea del mismo se da en ubicarse en un momento t y analizar la parcialidad de la secuencia de observaciones que determinan a O hacia atrás y hacia adelante, de allí proviene su nombre genérico.

4.2.2. Problema de decodificación. El problema de decodificación consiste en saber cuál es la secuencia de estados más probable del modelo λ que produce una determinada secuencia de observaciones $O = o_1, ..., o_T$. Para ello se tiene el algoritmo de *Viterbi*, el cual permite encontrar una secuencia de estados completa con la máxima verosimilitud.

De acuerdo a [10], el algoritmo se da en 4 pasos.

Algoritmo 1 Algoritmo de Viterbi

Paso 1: Inicialización
$\delta_i(i) = \pi_i b_i(o_1),\ 1 \leq i \leq N$
$\psi_i(i) = 0$
Paso 2: Recursión
Para $2 \leq t \leq T$, $1 \leq j \leq N$
$\delta_t(j) = max[\delta_{t-1}(i)a_{ij}]b_j(o_t)$
$\psi_t(j) = \arg \max_{1 \leq i \leq N}[\delta_{t-1}a_{ij}]$
Paso 3: Terminación $P^* = \max_{1 \leq i \leq N}[\delta_T(i)]$
$i_T^* = \arg \max_{1 \leq i \leq N}[\delta_T(i)]$
Paso 4: Trayecto de regreso (secuencia de estados)
Para $t = T-1, T-2, ..., 1$
$i_t^* = \psi_{t+1}(i_{t+1}^*)$

La variable $\delta_t(i) = \max_{q_1, q_2, ..., q_{t-1}} P\{q_1, q_2, ..., q_{t-1}, q_t = i, o_1, o_2, ..., o_{t-1}|\lambda\}$ brinda la probabilidad más alta de la secuencia de estados para la observación parcial, antes de llegar al tiempo t.

4.2.3. Problema de aprendizaje. El problema de aprendizaje busca responder a la interrogante siguiente, ¿cómo ajustar los parámetros de un modelo λ para maximizar la $P\{O|\lambda\}$? donde $O = o_1, ..., o_T$ es una secuencia de observaciones dada.

Para abordar y dar solución a este problema existen diferentes procesos para optimizar los parámetros. Sin embargo, los más comunes para los MOMs son dos criterios, el criterio de la máxima verosimilitud (ML) y el criterio de máxima información mutua (MMI) [7]. El primer criterio, ML, considera procesos de optimización como el algoritmo de *Baum-Welch* o métodos gradientes para localizar los parámetros que maximizan la función de verosimilitud. Mientras que el segundo criterio, MMI, considera métodos de entrenamiento discriminativos buscando minimizar la incertidumbre condicional.

5. Implementación de los MOMs

Considerando los aspectos teóricos preliminares, se plantea un problema extraído de [11] para mostrar un ejemplo del proceso de implementación de un MOM, previo

a los análisis que se harán para otros casos.

El problema supone un proceso de producción de artículos iguales, que se generan por diferentes máquinas (unidades de producción), las cuales son independientes. Particularmente, se considera un caso donde solamente se tienen 2 máquinas que producen los artículos, los cuales pueden ser producidos correctamente o con fallos. Además, las máquinas pueden estar en estado normal w_1 o en estado anormal, w_2. El modelo considera las siguientes hipótesis:

1. Mientras una máquina esté produciendo un artículo, el estado de la máquina no cambia.
2. Si una máquina está en el estado w_1, inmediatamente luego de producir un artículo puede pasar al estado w_2 con probabilidad p y permanecer en el estado w_1 con probabilidad $1-p$.
3. Si una máquina está en el estado w_2, permanecerá en él hasta que se realice el mantenimiento correctivo para pasar al estado w_1.
4. Un artículo producido puede ser clasificado como conforme o no conforme.
5. Cuando una máquina está en el estado w_i, la probabilidad de producir un artículo conforme es r_i y de producir un artículo no conforme es $1-r_i$, con $i=1,2$ y $r_1 > r_2$.

Con ello, se tiene que los estados ocultos del MOM corresponden a los estados de las máquinas, w_1 y w_2. Mientras que las señales (observaciones) se obtienen de una inspección del artículo, que se supone perfecta, para clasificarlo en conforme o no conforme.

Los autores plantean en forma general el modelo, considerando n máquinas independientes. De ello, consideran que la matriz de transición de probabilidad entre estados de la máquina esta dada por

$$A_{n+1} = \{a_{ij}\} \text{ donde } \; a_{ij} = C_{j-i}^{n-i+1}(1-p)^{n-j+1}p^{j-i}, \; (i,j=1,2,...,n+1),$$

y C_k^m es el número de combinaciones de k artículos de un total de m unidades siempre que $0 \leq k \leq m$ y cero en otro caso.

La matriz de distribución de probabilidad de los estados observables de n artículos producidos, uno de cada una de las n máquinas está dada por

$$B_{n+1} = \{b_i(k)\}, \text{ donde } \; b_i(k) = \begin{cases} \sum_{l=0}^{k} p(k-l,l), & k \leq i,\ k \leq n-i \\ \sum_{l=k-(n-i)}^{k} p(k-l,l), & k \leq i,\ k > n-i \\ \sum_{l=0}^{i} p(k-l,l), & k > i,\ k \leq n-i \\ \sum_{l=k-(n-i)}^{i} p(k-l,l), & k > i,\ k > n-i \end{cases}$$

Con $i,k = 0,1,2$, donde $p(k-l,l) = C_{k-l}^{n-i} r_1^{(n-i)-(k-l)}(1-r_1)^{k-l} \times C_l^i r_2^{i-l}(1-r_2)^l$

Para el caso particular donde $n=2$, se tienen las matrices

$$A_3 = \begin{bmatrix} (1-p)^2 & 2p(1-p) & p^2 \\ 0 & (1-p) & p \\ 0 & 0 & 1 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} r_1^2 & 2r_1(1-r_1) & (1-r_1)^2 \\ r_1 r_2 & (1-r_1)r_2 + (1-r_2)r_1 & (1-r_1)(1-r_2) \\ r_2^2 & 2r_2(1-r_2) & (1-r_2)^2 \end{bmatrix}$$

Tomando $r_1 = 0{,}95$, $r_2 = 0{,}05$, $p = 0{,}02$ y $\pi = (1, 0, 0)$ que implica que ambas máquinas están en estado normal al momento inicial. Considerando que las máquinas no tienen distinción, se generan tres estados para el sistema de producción, los cuales son
$S_1 = \{w_1, w_1\}$ que significa que ambas máquinas están en estado normal,
$S_2 = \{w_1, w_2\}$, una máquina esta en estado normal y la otra en estado anormal, y
$S_3 = \{w_2, w_2\}$, que indica que ambas máquinas están en estado anormal o de falla

De acuerdo a los estados planteados, solo se determina cuantas máquinas están fallando, sin indicar cual de las dos es la que falla. Con ello, se tienen tres estados para los dos artículos que se generan de las máquinas, uno de cada máquina.

$$v_1 = \{u_1, u_1\},\ v_2 = \{u_1, u_2\},\ v_3 = \{u_2, u_2\}$$

El modelo se representa en el siguiente diagrama

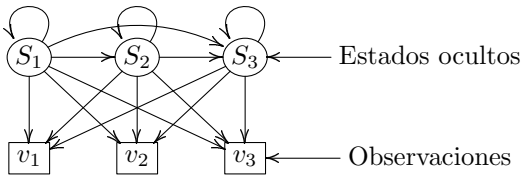


FIGURA 1. Diagrama del MOM

Considerando lo expuesto por [11] se generan 50 secuencias de 200 observaciones cada una a partir de los parámetros del MOM, dados por A_3, B_3 y π, lo cual se resume en el cuadro 1.

CUADRO 1. Resumen de resultados de simulación

		Exactos	Antes	Después
Primer deterioro	Frecuencia	22	13	15
	Media de pasos	0	4	2.93
	Desviación estándar	0	3.40	3.23
Segundo deterioro	Frecuencia	12	22	16
	Media de pasos	0	5.35	5.19
	Desviación estándar	0	7.77	3.75

De los resultados anteriores, solamente en dos casos no se predijo el cambio del estado S_2 al S_3. Desestimando esos dos casos para las medidas estadísticas, se obtienen resultado adecuados según Tai, Ching y Chan en [11]. Además, el primer deterioro lo predice mejor el modelo, esto porque a pesar que la diferencia de los pasos entre el primer deterioro y el segundo resulte igual al proceso real, la dependencia del segundo deterioro con el primero hace que se evidencie la discrepancia.

Por otro lado, si se busca distinguir la máquina que está fallando solo se debe ajustar la cantidad de estados ocultos, para luego ajustar la matriz de transición entre estados. En este caso se tendrían los siguientes estados:
$S_1 = (w_1, w_1)$ ambas máquinas están en estado normal.

$S_2 = (w_1, w_2)$ la primera máquina está en estado normal, mientras la segunda en estado anormal.
$S_3 = (w_2, w_1)$ la primera máquina está en estado anormal, mientras la segunda en estado normal.
$S_4 = (w_2, w_2)$ ambas máquinas están en estado anormal.

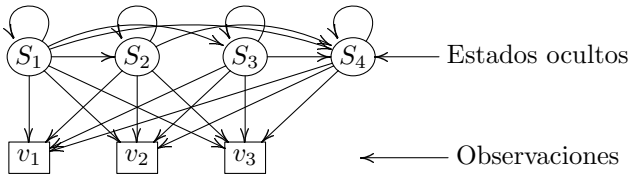


FIGURA 2. Diagrama del MOM con distinción de máquinas

Considerando que p'_i es la probabilidad de que la máquina i, que está en el estado w_1 permanezca en él, y p_i de que la máquina que está en el estado w_1, cambie al estado w_2, con $i = \{1, 2\}$ se tiene la matriz de transición de estados siguiente.

$$A = \begin{bmatrix} p'_1 p'_2 & p_1 p'_2 & p'_1 p_2 & p_1 p_2 \\ 0 & p'_1 & 0 & p_1 \\ 0 & 0 & p'_2 & p_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Los estados de salida serían los mismos tres del primer caso, lo que origina una matriz de emisión de 4×3, donde las filas serían los estados S_i para $i = 1, 2, 3, 4$, y las columnas los estados v_j donde $j = 1, 2, 3$. Definiendo a r_{ik} como la probabilidad de que la máquina i en el estado w_k produzca un artículo conforme y, r'_{ik} de que no produzca un artículo conforme, se tiene la matriz de emisión siguiente

$$B = \begin{bmatrix} r_{11}r_{21} & r'_{11}r_{21} + r_{11}r'_{21} & r'_{11}r'_{21} \\ r_{11}r_{22} & r'_{11}r_{22} + r_{11}r'_{22} & r'_{11}r'_{22} \\ r_{12}r_{21} & r'_{12}r_{21} + r_{12}r'_{21} & r'_{12}r'_{21} \\ r_{12}r_{22} & r'_{12}r_{22} + r_{12}r'_{22} & r'_{12}r'_{22} \end{bmatrix}$$

Considerando los valores para $p_1, p_2, r_{11}, r_{12}, r_{21}, r_{22}$ y π siguientes

$$p = \begin{bmatrix} 0{,}008 \\ 0{,}015 \end{bmatrix} \; r = \begin{bmatrix} 0{,}65 & 0{,}3 \\ 0{,}8 & 0{,}4 \end{bmatrix} \text{ y } \pi = [1, 0, 0, 0],$$

que implica que el sistema de producción comienza con las dos máquinas funcionando normalmente, se generan 200 muestras de 200 datos sintéticos para estimar en que pasos hay cambio de estados, e identificar la máquina que esta fallando. Los resultados se muestran en el cuadro 2, donde se tienen 156 procesos correctos, esto se refiere a que identifica los cambios de estado adecuados, es decir, si en la secuencia real la transición de estados es de S_1 a S_3 y luego de S_3 a S_4 , en la secuencia simulada resulta igual. Cuando un cambio en la secuencia aproximada no concuerda o, el segundo cambio no se efectúe, se consideran como casos incorrectos.

CUADRO 2. Resumen de resultados de simulación para caso con distinción de máquina

		Exactos	Antes	Después
Primer deterioro	Frecuencia	18	80	58
	Media de pasos	0	7.84	7.36
	Desviación estándar	0	10.73	8.34
Segundo deterioro	Frecuencia	27	59	70
	Media de pasos	0	7.02	13.43
	Desviación estándar	0	19.18	9.86

Las estadísticas de las 156 simulaciones que brindaron transiciones adecuadas, muestran datos apropiados conforme lo planteado en [11]. Sin embargo, se evidencia mayor dispersión que la obtenida con el primer enfoque. Por otro lado, respecto a ambos enfoques, resulta necesario explorar en función del contexto las técnicas apropiadas para estimar los parámetros del modelo, en este caso, las matrices de transición entre estados y de emisión. A pesar de ello, se replica la idea y se evidencia una estructura y estrategia base para problemas de estimación de fallas en un proceso de producción con cierta cantidad de máquinas.

Otros trabajos, como el planteado por Arpaia, et. al. en [1] proponen una metodología interesante para determinar los estados ocultos de un MOM considerando datos escasos y con falta de sincronización. La no sincronización se refiere a que los sensores que permiten obtener los datos no los brindan en el mismo instante, lo que conlleva a tener que hacer un proceso de filtrado.

El modelo que plantean supone que los datos se obtienen en una condición operativa estacionaria, es decir, que las medidas no se adquieren durante las transiciones de la máquina de un régimen a otro. En este caso los autores entrenan el modelo para obtener las matrices de transición con los datos filtrados, donde inicialmente en el modelo no se conoce el número de estados ocultos. Por lo que emplean el algoritmo de Baum-Welch para encontrar el número de estados que maximiza una función de verosimilitud, esto mediante el algoritmo de Expectación-Máxima (EM).

Luego de ello, plantean el proceso de validación donde consideran una prueba de bondad de ajuste entre la muestra de verosimilitud obtenida en el entrenamiento y la obtenida por la aplicación de la secuencia de prueba. Se detecta un fallo cuando se encuentra una distribución diferente para las dos muestras.

El fenómeno estudiado consideraba como caso de estudio los compresores de tornillo de un sistema criogénico de un gran colisionador de hadrones. En la aplicación del modelo, que consistía en una máquina específica de fluidos, no se poseían datos de fallas, por lo que se planteó alterar de forma sintética los datos para indicar una falla, mostrando así los rangos de valores de las características que serían indicativos de una falla en la máquina.

En ese sentido, la estrategia que plantearon logró establecer rangos de las señales donde supondría una falla al compresor de tornillo.

Considerando lo anterior, se plantea el algoritmo de detección de fallas en máquinas de fluidos. Para ello, se necesitan definir tres funciones, una que realice el pre procesamiento de los datos, otra para el entrenamiento del modelo y la última para la prueba del modelo, la cual permite determinar las anomalías según lo planteado por los autores.

La función de Preprocesameinto de secuencias dada en el algoritmo 2 realiza lo siguiente:

1. Filtrado: Se eliminan las secuencias que no están en el período de funcionamiento (ON).
2. Relleno: Se replica la última medición disponible para completar la longitud de las secuencias filtradas.
3. Decimación o diezmado: Se selecciona una observación cada N, donde N es el factor de decimación.
4. ACP: Se utiliza Análisis de Componentes Principales (ACP) para extraer los componentes principales de las secuencias diezmadas. Con el ACP se logra una reducción de dimensionalidad que ayuda a simplificar los datos mientras mantiene su información relevante.

Algoritmo 2 Preprocesamiento de secuencias

Entrada: *sequences*: Las secuencias a preprocesar.
Salida: Las secuencias preprocesadas.

```
filtered_sequences ← Lista vacía
for sequence en sequences do
    if sequence[0] = 1 then
        filtered_sequences.agregar(sequence)
    end if
end for
filled_sequences ← Lista vacía
for sequence en filtered_sequences do
    filled_sequence ← Un arreglo de ceros de la misma longitud que sequence
    filled_sequence[: len(sequence)] ← sequence
    filled_sequences.agregar(filled_sequence)
end for
decimated_sequences ← Lista vacía
for sequence en filled_sequences do
    decimated_sequence ← sequence[:: 2]
    decimated_sequences.agregar(decimated_sequence)
end for
Calcular la matriz de componentes principales principal_components usando
PCA con n_components = 3 y las secuencias en decimated_sequences.
    return principal_components
```

Por otro lado la función de entrenamiento de MOMs, mostrada en el algoritmo 3, permite entrenar los MOMs utilizando el algoritmo de aprendizaje no supervisado

llamado KMeans y los datos preprocesados obtenidos en la etapa anterior. Para cada secuencia, se aplica el algoritmo KMeans para agrupar las secuencias en diferentes clusters. Luego, se crea un MOM y se ajusta a los datos de los componentes principales de la secuencia. Cada modelo MOM representa un grupo de secuencias similares.

Algoritmo 3 Entrenamiento de MOMs

Entrada: *sequences*: Las secuencias para entrenar los modelos HMM.
Salida: Los modelos HMM entrenados.

```
hmm_models ← Lista vacía
for sequence en sequences do
    cluster ← KMeans(principal_components)
    hmm ← GaussianHMM(n_components=3)
    sequence ← np.reshape(sequence, (1, -1))
    hmm.fit(principal_components)
    hmm_models.append(hmm)
end for
    return hmm_models
```

Finalmente, la función prueba de MOMs en secuencias, dada en el algoritmo 4, prueba las secuencias de datos procesados para detectar fallas. En este caso, para cada secuencia se calcula la probabilidad logarítmica del MOM asociado a esa secuencia. Si la probabilidad logarítmica es menor que el umbral especificado, la secuencia se considera como rechazada, lo que significa que el MOM no es capaz de explicar bien esa secuencia y se considera como anómala o con falla.

Algoritmo 4 Prueba de MOMS en secuencias

Entrada: *models*: Los MOMs a probar.
Entrada: *sequences*: Las secuencias para probar los MOMs.
Entrada: *threshold*: Umbral para rechazar secuencias basado en la probabilidad MOM.
Salida: El número de secuencias rechazadas.

```
rejected_sequences ← 0
for sequence en sequences do
    for model en models do
        Reformar la secuencia a matriz 2D antes de predecir
        sequence_2d ← sequence.reshape(1, -1)
        log_prob ← model.score(sequence_2d)
        if log_prob < log(threshold) then
            rejected_sequences ← rejected_sequences + 1
            break
        end if
    end for
end for
    return rejected_sequences
```

Con los algoritmos anteriores, se generan 20 muestras sintéticas de 100 secuencias con 100 elementos cada una para implementar las ideas dadas en [1], los resultados se muestran en el cuadro 3.

CUADRO 3. Secuencias entrenadas y secuencias con fallas

Entrenadas	44	50	55	38	49	43	51	44	58	52
Rechazadas	42	48	35	36	47	41	46	42	52	50
Entrenadas	38	46	47	58	50	53	53	53	51	55
Rechazadas	32	37	43	57	49	49	52	52	48	53

Se puede observar que la cantidad de secuencias rechazadas, que indican fallos en el dispositivo, es significativamente alta en comparación con la cantidad de secuencias que son exitosamente entrenadas después del proceso de preprocesamiento. Esta disparidad sugiere que el modelo es altamente sensible en la detección de fallas. Esta alta sensibilidad se traduce en una capacidad efectiva para identificar situaciones anómalas en la máquina de fluidos.

6. Conclusiones

Conforme a lo abordado anteriormente se concluye que:

1. Los MOMs resultan adecuados para lograr detectar fallas en un proceso. Sin embargo, se debe analizar con detalle la selección de los estados ocultos y la obtención de la información necesaria para entrenar el modelo. Además, la detección de fallas corresponde a conocer la secuencia de estados ocultos más probable, por lo que el algoritmo de Viterbi, o las mejoras al mismo, se deben explorar y analizar para lograr no solo una detección adecuada, sino también eficiente.

2. La determinación de estados ocultos depende de las hipótesis o especificidad que se requiere en cuanto al fenómeno. Sin embargo, cuando se trabaja con grandes cantidades de estados, conviene establecer una técnica para tomar los estados más significativos, sin que se pierda precisión y eficacia en el modelo. Esto se puede lograr implementando algunas técnicas como ser el Análisis de Componentes Principales (ACP).

3. Los MOMs logran identificar anomalías al analizar exclusivamente datos en los que no se hayan registrado fallos previos. Esta particularidad constituye una ventaja significativa, ya que, según los resultados obtenidos, su capacidad para detectar cambios sutiles se revela beneficiosa para estimar rasgos que denoten una posible falla. No obstante, es esencial evaluar el contexto de aplicación y explorar diversas estrategias para el preprocesamiento de los datos, la fase de entrenamiento y la selección adecuada del umbral que permita identificar secuencias anómalas de manera precisa.

Referencias

1. P. Arpaia, U. Cesaro, M. Chadli, Hervé Coppier, L De Vito, A Esposito, F. Gargiulo, and M. Pezzetti, *Fault detection on fluid machinery using hidden markov models*, Measurement **151** (2020), 107126.

2. P. Bickel, P. Diggle, S. Fienberg, and U. Gather, *Springer series in statistics*, (2005).
3. Jakub Michal Bilski and Agnieszka Jastrzebska, *Fuzzy cognitive maps and hidden markov models: Comparative analysis of efficiency within the confines of the time series classification task*, arXiv preprint arXiv:2204.13455 (2022).
4. Luini Leonardo Hurtado Cortés, Edwin Villarreal-López, and Luís Villarreal-López, *Detección y diagnóstico de fallas mediante técnicas de inteligencia artificial, un estado del arte*, Dyna **83** (2016), no. 199, 19–28.
5. Przemyslaw Dymarski, *Hidden markov models: Theory and applications*, BoD–Books on Demand, 2011.
6. Abdelmalek Kouadri, Mansour Hajji, Mohamed-Faouzi Harkat, Kamaleldin Abodayeh, Majdi Mansouri, Hazem Nounou, and Mohamed Nounou, *Hidden markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems*, Renewable Energy **150** (2020), 598–606.
7. Guy Leonard Kouemou and Dr Przemyslaw Dymarski, *History and theoretical basics of hidden markov models*, Hidden Markov models, theory and applications **1** (2011).
8. Bhavya Mor, Sunita Garhwal, and Ajay Kumar, *A systematic review of hidden markov models and their applications*, Archives of computational methods in engineering **28** (2021), no. 3, 1429–1448.
9. Antonio Muñoz, José Rodríguez Herrerías, and José María Martínez Val, *La seguridad industrial: Fundamentos y aplicaciones*, Fundación para el Fomento de la Innovación industrial, 2005.
10. Lawrence Rabiner and Biinghwang Juang, *An introduction to hidden markov models*, ieee assp magazine **3** (1986), no. 1, 4–16.
11. Allen H. Tai, Wai-Ki Ching, and Ling-Yau Chan, *Detection of machine failure: Hidden markov model approach*, Computers & Industrial Engineering **57** (2009), no. 2, 608–619.
12. Zhenyu Wu, Hao Luo, Yunong Yang, Peng Lv, Xinning Zhu, Yang Ji, and Bian Wu, *K-pdm: Kpi-oriented machinery deterioration estimation framework for predictive maintenance using cluster-based hidden markov model*, IEEE Access **6** (2018), 41676–41687.
13. Shun-Zheng Yu, *Hidden semi-markov models: theory, algorithms and applications*, Morgan Kaufmann, 2015.

Email address: david.ordonez@unah.edu.hn

ANÁLISIS DE UNA FAMILIA DE CURVAS ELÍPTICAS EN CRIPTOGRAFÍA

HÉCTOR JAVIER FLORES ORDÓÑEZ

This paper is dedicated to my family.

Abstract. Elliptic curve cryptography has proven to be a valuable tool in safeguarding sensitive information in digital environments. In this article, we delve into a specific family of elliptic curves used in cryptography and explore why they are chosen to ensure the confidentiality, integrity, and authenticity of data. Specifically, we focus on elliptic curves over finite fields, such as prime fields. We explore the fundamental properties of these curves, the selection of security parameters, and examine their performance and computational efficiency compared to other cryptographic schemes.

Resumen. La criptografía de curvas elípticas ha demostrado ser una herramienta valiosa en la protección de la información sensible en entornos di gitales. En este artículo, se analiza en detalle una familia de curvas elípticas utilizadas en criptografía y se explora por qué son elegidas para asegurar la confidencialidad, integridad y autenticidad de los datos. En particular, nos enfocamos en las curvas elípticas sobre cuerpos finitos. Exploramos las propiedades fundamentales de estas curvas, la elección de parámetros de seguridad, y examinamos su rendimiento y eficiencia computacional en comparación con el esquema RSA.

Fecha: August 30, 2023.
Palabras y frases clave. Criptografía, Curvas Elípticas, Campos Finitos.

1. Introducción

Históricamente se ha tenido la necesidad de compartir información, en muchos casos esta información es confidencial y debe mantenerse de dicha forma entre emisor y receptor, este problema ha motivado diferentes métodos de encriptación, entre los cuales se pueden mencionar dos tipos: *encriptación simétrica* y *encriptación asimétrica*, la primera de estas se refiere a que tanto el emisor como el receptor conocen una clave única que puede encriptar y desencriptar un mensaje [1], esto parece muy útil y eficaz, siempre y cuando el emisor y el receptor puedan coincidir en un lugar para establecer dicha clave, de lo contrario, cualquiera que pueda interceptar el mensaje, podría de la misma forma interceptar la clave.

El problema antes mencionado, se resuelve con la encriptación asimétrica introducida por Withfield Diffie y Martin Hellman en 1975 [3], bajo el concepto de encriptación de clave pública, de la cual existen diversas variantes [2].

La criptografía desempeña un papel crucial en la protección de la información confidencial en la era digital. A medida que los avances tecnológicos continúan, es imperativo desarrollar sistemas criptográficos sólidos y eficientes para garantizar la seguridad de la comunicación y el almacenamiento de datos sensibles. En este contexto, la criptografía de curvas elípticas ha emergido como una herramienta poderosa y ampliamente utilizada.

Las curvas elípticas, que son una rama de la geometría algebraica, han demostrado ser una base sólida para la criptografía moderna. Su estructura matemática única y sus propiedades especiales las hacen adecuadas para aplicaciones criptográficas, ofreciendo una mayor seguridad con claves más cortas en comparación con otros algoritmos criptográficos tradicionales [4], esto se traduce en menor consumo de recursos en almacenamiento de claves, procesamiento en verificación de claves y ancho de banda en la transmisión de la información.

En este artículo, nos enfocaremos en explorar una familia específica de curvas elípticas utilizadas en criptografía y examinar por qué estas curvas han sido elegidas para garantizar la confidencialidad, integridad y autenticidad de los datos. Analizaremos las propiedades fundamentales de estas curvas y cómo se utilizan en aplicaciones criptográficas, como el cifrado de clave pública y el esquema de firma digital [5], de la misma forma también se presenta el problema del logaritmo discreto, que es la base de la seguridad criptográfica de estas curvas. Finalmente, se compararán computacionalmente frente a los métodos tradicionales de encriptación.

Específicamente, se trabajará con curvas elípticas sobre cuerpos finitos , es fundamental seleccionar cuidadosamente la familia de curvas y los parámetros adecuados según los requisitos de seguridad y eficiencia específicos del sistema.

1.1. **Justificación.** Actualmente se depende constantemente de la transferencia de información, ya sea para comunicarse a través de sistemas de mensajería instantánea, para realizar compras en línea, ejecutar transferencias bancarias, etc. Al momento de realizar estos procesos se desea que sean ejecutados de forma confidencial, emisor-receptor, es en este punto donde surge la necesidad de tener un método que permita compartir información de forma segura, dando nacimiento a los sistemas de encriptación.

Actualmente uno de los métodos más populares de encriptación es el RSA [7], nombrado así en honor a sus desarrolladores, R.L. Rivest, A. Shamir y L. Adleman [6]. Este método consiste en la factorización de números enteros [6], pero presenta ciertas desventajas frente a los algoritmos basados en encripatación de curvas elípticas, como la mayor longitud en las claves, marcando una diferencia en tiempo de ejecución, seguridad y potencia [8].

Un modelo matemático que permita desarrollar algoritmos más eficientes representa un avance para poder consumir menos recursos y bridar mayor facilidad al momento de implementar estos conocimientos en nuevas tecnologías que permitan la transferencia de información de forma segura, generando aportes de importancia en la ciencia y siendo un campo fértil de investigación debido al desarrollo exponencial de la ciencia computacional.

Para poder desarrollar con claridad este trabajo es necesario que se conozca un poco de la evolución de los criptosistemas, así como las bases matemáticas que los sustentan, debido a ello en la siguiente sección se muestran los preliminares necesarios.

2. Antecedentes

2.1. **Encriptación Asimétrica.** Para analizar este concepto considere que se crea un canal de comunicación, el cual necesita una llave para abrirse y otra para cerrarse, de modo que, la llave que abre el canal no puede cerrarlo y la llave que cierra el canal no puede abrirlo.

Ahora dígase que el individuo A crea el canal y se queda con la llave de apertura, luego envía al individuo B las instrucciones del canal y la llave que cierra dicho canal. Nótese que cualquiera que intercepte las instrucciones y la llave, no podrá abrir dicho canal, de modo que cuando el individuo B redacte su mensaje y cierre el canal, el único que podrá abrirlo será el individuo A.

Withfield Diffie y Martin Hellman en 1975 [3] introducen el concepto de *encriptación de llave pública*, resolviendo así los inconvenientes mostrados en la encriptación simétrica, ya que cada individuo debe elegir un par de claves, una *clave pública* y una *clave privada*, esta clave pública debe ser compartida con el receptor sin temor a que sea interceptada, ya que conocerla no es suficiente para poder conocer la clave privada del emisor.

Actualmente existen diversos métodos de clave pública, estos se basan en la idea de funciones con una trampilla, funciones que son fáciles de tratar en una dirección, pero su inversa toma exponencialmente más tiempo cuando no se tiene la clave de desencriptación.

2.2. **RSA.** Desde su innovación en el año 1977 es considerado como uno de los criptosistemas más seguros [7]. El RSA basa su seguridad en el problema de factorización de números enteros. A continuación, se resumen los conceptos básicos del algoritmo RSA descritos en [6].

2.2.1. *Etapas Del Algoritmo.* El algoritmo hace uso de la función de Euler $\varphi(n)$, esta función se define para un entero positivo n, como la cantidad de enteros positivos menores a n y coprimos con n.

Algorithm 1: Algoritmo para generación de claves en el método RSA

Data: p, q primos tales que $p \neq q$; ▷ `Estos valores deben ser secretos`

Result: $(n, e), (n, d)$; ▷ `Clave pública y clave privada respectivamente`

$n \leftarrow pq$;

$\varphi(n) \leftarrow (p-1)(q-1)$; ▷ `Función de Euler`

Seleccionar un entero d tal que $\text{MCD}(d, \varphi(n)) = 1$;

Calcular el valor e tal que $ed \equiv 1 \text{mod}(\varphi(n))$; ▷ $d \equiv e^{-1} \texttt{mod}(\varphi(n))$

Compartir (n, e) y guardar en secreto (n, d);

Algorithm 2: Algoritmo para encriptación de claves en el método RSA

Data: mensaje x , clave pública (n, e)
Result: Mensaje encriptado y
$y \leftarrow x^e \text{mod}(n)$; ▷ `Encriptación del mensaje`
Enviar el mensaje encriptado ;

Algorithm 3: Algoritmo para desencriptación de claves en el método RSA

Data: mensaje encriptado y, clave privada (n, d)
Result: Mensaje x
$x \leftarrow y^d \text{mod}(n)$; ▷ `Desencriptación del mensaje`

2.3. **Criptografía De Curvas Elípticas.** La criptografía de curvas elípticas requiere mayores conocimientos matemáticos que el RSA, de modo que antes de describir el método se deben mencionar algunos conceptos relacionados con este problema. Los conceptos se presentan en una forma básica, ya que la idea es comprender el funcionamiento, pero en el próximo capítulo se analiza con mayor detalle como son aplicadas a la criptografía.

Las curvas elípticas fueron introducidas en criptografía por Neal Koblitz en [10] y Victor S. Miller en [11]. Ambas propuestas fueron pioneras y marcaron el inicio del estudio y desarrollo de la criptografía de curva elíptica. Desde entonces, ECC, por sus siglas en inglés, ha ganado popularidad debido a su eficiencia, seguridad y la capacidad de proporcionar una seguridad similar con claves más cortas en comparación con otros esquemas criptográficos tradicionales, como RSA y sistemas basados en logaritmos discretos. ECC se ha convertido en un componente esencial de muchos protocolos criptográficos y sistemas de seguridad utilizados en la actualidad.

A diferencia de RSA que basa su seguridad en la factorización de números $n = pq$, donde p y q son las claves y son primos muy grandes, actualmente sugeridos de 2048 bits, ECC utiliza como claves puntos sobre las curvas elípticas y basa su seguridad en el problema del *logaritmo discreto* del cual se hablará más adelante.

3. Conceptos Básicos Sobre Curvas Elípticas

Una curva elíptica E es la gráfica de una ecuación

$$E: \qquad y^2 = x^3 + ax + b$$

donde x, y, a y b pertenecen a algún campo K, que podrían ser $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ o $\mathbb{F}_p$ donde p es un número primo y con $4a^3 + 27b^2 \neq 0$, está última condición garantiza la suavidad en la curva.

3.1. Álgebra Sobre Curvas Elípticas.

Comenzamos con $K = \mathbb{R}$ ya que nos ayuda a tener una mejor interpretación visual del álgebra.

Definición: Sean P y Q dos puntos distintos sobre una curva elíptica, la recta que une estos dos puntos cortará la curva en un tercer punto que llamaremos R, luego la reflexión de este punto en el **eje x** se denominará $P + Q$ y es ilustrado en la siguiente figura.

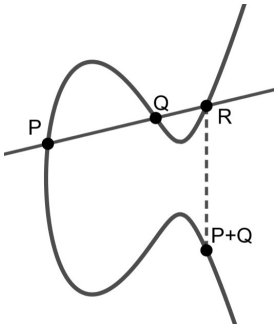


Figure 1. Representación gráfica de suma $P + Q$ donde $P \neq Q$ elaborada en Geogebra.

Si $P = Q$, se traza una recta tangente a la curva en P que corta la curva en un segundo punto, R, el resultado de $P + Q = P + P$ es el reflejo en el **eje x** del punto R.

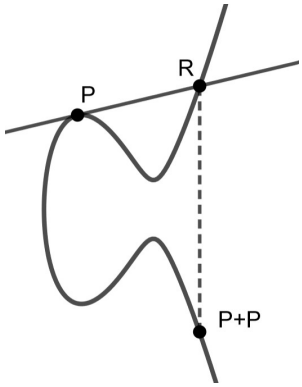


Figure 2. Representación gráfica de suma $P + P$ elaborada en Geogebra.

Es importante incluir el punto infinito ∞, este no pertenece a la curva, pero desempeña como elemento identidad con respecto a la suma, es decir, si $P = \infty$, entonces $P + Q = Q + P = Q$. Se establece que las rectas que pasan por ∞ son rectas verticales, de modo que una recta vertical que pasa por (x, y) corta la curva en $(x, -y)$ y al reflejarse en el eje x, regresa a (x, y).

De la geometría anteriormente mencionada se puede deducir que para todo elemento $P = (x, y)$ existe un elemento $-P = (x, -y)$, tal que $P + (-P) = \infty$ y finalmente, también la propiedad asociativa, de modo que tenemos un grupo algebraico que también es abeliano.

De la geometría mostrada es posible calcular el valor de $P + Q$ encontrando la ecuación de la recta que pasa por los puntos P y Q y sustituyendo en la ecuación de la curva, de modo que debemos encontrar las raíces de la nueva ecuación que es de grado 3, pero que ya conocemos dos de esas raíces, P y Q, por lo tanto solo es necesario calcular el valor de R y reflejarlo en el eje x.

Sean $P = (x_1, y_1)$, $Q = (x_2, y_2)$ puntos diferentes y sea $y = \alpha x + \beta$ la ecuación de la recta que los une, entonces

$$\alpha = \frac{y_2 - y_1}{x_2 - x_1} \qquad \text{y} \qquad \beta = y_1 - \alpha x_1$$

y la ecuación de la recta es

$$y = \frac{y_2 - y_1}{x_2 - x_1} x + y_1 - \alpha x_1$$

Ahora sustituyendo en la ecuación de la curva, obtenemos

$$(\alpha x + \beta)^2 = x^3 + ax + b$$

es decir, debemos encontrar las raíces de

$$0 = x^3 - \alpha^2 x^2 + (a - 2\alpha\beta)x + (b - \beta^2)$$

donde sabemos que x_1 y x_2 son raíces, también sabiendo que en un polinomio cuyo coeficiente principal es 1, la suma de las raíces es igual al negativo del coeficiente de la segunda potencia más alta, es decir

$$x_3 + x_2 + x_1 = \alpha^2$$

y se obtiene el resultado de que $x_3 = \alpha^2 - x_2 - x_1$.

Al saber que estos puntos pertenecen a la recta $y = \alpha x + \beta$ los podemos enlistar como:

* $P = (x_1, \alpha x_1 + \beta)$
* $Q = (x_2, \alpha x_2 + \beta)$
* $P + Q = (x_3, -(\alpha x_3 + \beta))$

y finalmente,

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_2 - x_1$$

$$-(\alpha x_3 + \beta) = y_3 = -y_1 + \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x_1 - x_3)$$

Por otra parte si $P = Q$, el análisis es el mismo, con la diferencia de que $\alpha = \dfrac{dx}{dy}$ en P, obteniendo

$$x_3 = \left(\frac{3x_1^2 + a}{2y_1}\right)^2 - 2x_1$$
$$y_3 = -y_1 + \left(\frac{3x_1^2 + a}{2y_1}\right)(x_1 - x_3)$$

Todo lo anterior nos ayuda a garantizar que la suma sobre curvas elípticas está bien definida y a tener un método para realizar dicha suma sobre el campo $K = \mathbb{R}$, pero que también nos da un punto de partida para trasladarnos a campos finitos.

3.2. **Curvas Elípticas Sobre Campos Finitos.** Si bien es posible definir curvas elípticas sobre los números reales o los números complejos, para fines criptográficos es conveniente trabajar sobre campos con números enteros debido al problema de redondeo al momento de hacer operaciones en una computadora, específicamente trabajaremos con $\mathbb{F}_p$, donde p es un número primo y al conjunto de puntos de la curva los denotamos con $\mathrm{E}(\mathbb{F}_p)$.

Cabe destacar que la gráfica de curvas elípticas sobre estos campos finitos ya no conciden con la mostrada para los reales, pero sigue siendo un grupo abeliano.

Ejemplo 1: Si se considera E sobre $\mathbb{F}_7$ mediante la ecuación $y^2 = x^3 + 2x + 4$, los puntos de la curva son

$$E(\mathbb{F}_7) = \{\infty, (0,2), (0,5), (1,0), (2,3), (2,4), (3,3), (3,4), (6,1), (6,6)\}.$$

La forma de sumar puntos en $\mathbb{F}_p$ sigue siendo como la descrita para $\mathbb{R}$, pero haciendo la observación que una expresión racional como a/b debemos tratarla como ab^{-1}, donde $bb^{-1} \equiv 1 \bmod p$.

Veamos la suma de los puntos $(2,3)$ y $(6,1)$ sobre la curva antes mencionada en $\mathbb{F}_p$

$$\alpha \equiv \frac{1-3}{6-2} \equiv \frac{2}{-4} \equiv 2 * (-4)^{-1} \equiv 2 * 5 \equiv 6 \bmod 7$$

Luego

$$x_3 = 6^2 - 6 - 2 = 0$$
$$y_3 = -3 + 6(2 - 0) = 2$$

Por lo tanto $(2,3) + (6,1) = (0,2)$.

Una vez conocemos el álgebra de las curvas elípticas sobre campos finitos, es necesario discutir como utilizarlas para encriptar un mensaje. La seguridad del esquema se basa en el problema del logaritmo discreto y su variante aplicada a curvas elípticas, estos últimos conceptos se describen a continuación.

3.3. **Problema Del Logaritmo Discreto (DLP).** Según [9], para definir el DLP se considera un grupo abeliano finito $\mathbb{F}$ y sea $f \in \mathbb{F}$ de orden n en $\mathbb{F}$, es decir $f^n = e$ donde e es el elemento identidad de $\mathbb{F}$. Dado a que pertenece al grupo cíclico generado por f, $\langle f \rangle \subset \mathbb{F}$, se define el logaritmo discreto de a en base f, como el menor entero k, tal que $f^k = a$, es decir,

$$\log_f a = k \longleftrightarrow f^k = a. \tag{3.1}$$

Nótese que el DLP es la función inversa a la exponenciación modular, donde la exponenciación es una función considerablemente más fácil de calcular que su inversa gracias al algoritmo extendido de Euler [9], este es el punto en el que se basa la seguridad de diversos algoritmos de encriptación [5].

3.4. **El Problema Del Logaritmo Discreto En Curvas Elípticas (ECDLP).** En resumen, se puede decir que es el DLP aplicado al grupo abeliano definido por el conjunto de puntos sobre una curva elíptica en un campo finito y la operación suma antes mencionada, como lo define [5]. Sea E una curva elíptica definida sobre un campo finito $\mathbb{F}_p$ y sea $P \in E(\mathbb{F}_p)$ un punto de orden n, es decir, $nP = \infty$. A este punto P le llamaremos *generador* del grupo cofactor G.

Ahora consideremos un punto $Q = kP = P + P + ... + P$ para algún entero k. El problema radica en encontrar el valor de k a partir de Q y P.

4. Encriptación En Curvas Elípticas

En este punto ya es posible desarrollar el esquema de encriptación sobre curvas elípticas. Este proceso consiste en asignar el mensaje a un punto en la curva, luego que el emisor encripte este punto y finalmente el receptor lo decodifica. Estos tres pasos se describen a continuación.

4.1. **Incrustación Del Mensaje A La Curva.** El esquema para encriptar un mensaje m mediante curvas elípticas comienza con asignar la coordenada x de un punto de la curva a este mensaje m. Como se puede ver en el ejemplo 1, la curva está definida sobre $\mathbb{F}_7$, cuyos elementos son $\{0, 1, 2, 3, 4, 5, 6\}$, nótese que m no podría tomar los valores $\{4, 5\}$ puesto que no existe y tal que $y^2 \equiv m^3 + 2m + 4$, de modo que asignar una coordenada x a m no es tan sencillo si trabajamos sobre un campo finito.

Por otra parte, sabiendo que en un campo $\mathbb{F}_p$, con p primo, aproximadamente la mitad de los elementos serán cuadrados perfectos, es posible utilizar un método probabilístico, determinado por [2] para incrustar m a la curva con una probabilidad de falla aceptable.

Supongamos nuestro mensaje m está escrito como un número entero y lo buscamos incrustar en la coordenada x de un punto sobre la curva E, la probabilidad de que $m^3 + am + b$ sea un cuadrado perfecto es de $1/2$, entonces consideremos un entero grande K, tal que la probabilidad de falla al momento de incrustar el mensaje sea 2^{-K}. Supóngase que $(m + 1)K < p$. Tomando $x = mK + j$, para $j = 1, 2, ..., K - 1$, se comienza a evaluar $y^2 = f(x) = x^3 + ax + b$ hasta que $f(x)$ sea un cuadrado perfecto, si se encuentra y, entonces el mensaje incrustado será

$P_m = (x, y)$, si j recorre todos sus posibles valores y no se encuentra y entonces se ha fallado en la incrustación y se deben buscar otros parámetros.

Finalmente, si $x = mK + j$, es fácil ver que es posible recuperar m a partir de

$$m = \lfloor x/K \rfloor$$

donde $\lfloor x/K \rfloor$ denota el mayor entero menor que o igual a x/K.

Cabe destacar que la incrustación del mensaje no es la encriptación, el proceso de encriptación y desencriptación que se mostrará corresponde al esquema ElGamal elíptico [12], existen otros esquemas para este proceso sobre curvas elípticas que pueden consultarse en [2].

4.2. **Encriptación En ECC.** Ahora supongamos que Alice desea compartir de forma segura el mensaje m con Bob, también supongamos que este mensaje ya fue incrustado a la curva y ahora es el punto P_m. A diferencia del esquema RSA, las claves públicas son ligeramente más complicadas, primero deben definirse las claves públicas del canal, en este caso, corresponden al campo finito $\mathbb{F}_p$, la curva elíptica E y el punto generador P, luego las claves públicas personales de Alice y Bob, e_A y e_B, respectivamente y finalmente las claves privadas d_A de Alice y d_B de Bob.

Cuando Alice envía el mensaje P_m a Bob, lo encripta enviando los puntos

$$P_1 = d_A P \qquad \text{y} \qquad P_2 = P_m + d_A(e_B P)$$

Nótese que para que Bob o un tercero conozcan la clave privada d_A de Alice, deben resolver el ECDLP en P_1. Por otra parte, notar que P_2 es la suma, que está bien definida, de dos puntos sobre la curva, de modo que Alice ha enviado a Bob dos puntos que están sobre la curva E.

4.3. **Desencriptación En ECC.** En este momento Bob ha recibido el mensaje y debe desencriptarlo, el proceso es una operación bastante sencilla, que como podremos ver consume menos recursos que desencriptar un mensaje cifrado mediante el esquema RSA.

Ahora Bob, multiplica P_1 con su clave privada y lo resta de P_2 para obtener

$$P_2 - e_B P_1 = P_m + d_A(e_B P) - e_B(d_A P) = P_m + d_A(e_B P) - d_A(e_B P) = P_m$$

5. Resultados

A continuación mostramos una tabla comparativa sobre los tiempos que toma romper la seguridad del esquema RSA y ECC, cabe destacar que se han ejecutado con parámetros débiles para poder notar las diferencias en los tiempos, aunque los parámetros elegidos para el RSA son más robustos y el método es por fuerza bruta y para el ECC ElGamal se ha usado un número primo débil, puesto que el orden del grupo generado por P es $2 * 5 * 11 * 22303 * 36209 * 196539307$, el hecho de tener factores primos pequeños lo hace vulnerable y fácil de atacar mediante Pohlig-Hellman [13].

Aunque se han escogido parámetros débiles para el ECDLP y un método más eficiente que la fuerza bruta, el tiempo en romper el ECDLP es mayor que factorizar en el RSA, también hay que mencionar que el tiempo crece de forma exponencial a medida que se agregan más bits a las claves, en [14] se puede indagar más sobre las debilidades que pueden presentar algunas curvas elípticas si los parámetros no se eligen con detenimiento.

TABLE 1. Tabla comparativa de los tiempos que toma romper el esquema RSA y ECC.

ECC ElGamal	
Parámetros	Tiempo del ECDLP para $Q = kP$
$p = 17459102747413984477$	2.2308
k : Aleatorio; $10^{17} < k < p$, entre 60 y 64 bits	2.9218
$P_x = 15579091807671783999$	1.3540
$P_y = 4313814846862507155$	
$y = x^3 + 2x + 3$	2.2259
Algoritmo Pohlig-Hellman	1.6646
RSA	
Parámetros	Tiempo en factorizar $n = pq$
p : Aleatorio de 64 bits	0.6773
q : Aleatorio de 64 bits	0.7385
Algoritmo de fuerza bruta	0.4077
	1.6777
	0.8312

Fuente: Elaboración propia.

Los tiempos han sido medidos utilizando el software *SageMath*.

6. Conclusiones

La criptografía de curvas elípticas ha demostrado ser una herramienta valiosa para proteger información sensible en entornos digitales. Su eficiencia, seguridad y la capacidad de proporcionar una seguridad similar con claves más cortas en comparación con otros esquemas criptográficos tradicionales, como RSA, la hacen especialmente atractiva.

La criptografía de curvas elípticas se basa en propiedades matemáticas de las curvas elípticas sobre campos finitos. La suma de puntos en una curva elíptica forma un grupo abeliano, lo que es esencial para garantizar la seguridad y confiabilidad de los algoritmos criptográficos basados en estas curvas.

La seguridad de la criptografía de curvas elípticas se basa en la dificultad de resolver el problema del logaritmo discreto en el grupo formado por los puntos de la curva elíptica. Este problema, también conocido como ECDLP, implica encontrar el valor de k en la ecuación $Q = kP$, donde P es un punto generador y Q es otro punto sobre la curva.

Las curvas elípticas ofrecen ventajas significativas en términos de eficiencia computacional, lo que se traduce en un menor consumo de recursos y menor tiempo de ejecución en comparación con otros esquemas criptográficos, como RSA. Además, a pesar de que las claves utilizadas son más cortas, proporcionan un nivel de seguridad comparable.

La incrustación del mensaje en la curva es un paso importante para la encriptación en ECC. Se utiliza un método probabilístico para asignar el mensaje m a una coordenada x de un punto sobre la curva. Este proceso permite que el mensaje sea tratado como un punto en la curva elíptica, lo que facilita la encriptación.

El esquema de encriptación y desencriptación utilizado en ECC se basa en ElGamal elíptico. Este esquema ofrece un nivel de seguridad sólido y una operación eficiente para desencriptar el mensaje recibido.

En general, la criptografía de curvas elípticas representa un área importante de estudio e investigación en la ciencia computacional, y su aplicación tiene un impacto significativo en la seguridad de la información en la era digital. Con su eficiencia y sólidas bases matemáticas, esta tecnología seguirá desempeñando un papel clave en la protección de datos sensibles en diversos campos y aplicaciones.

7. Trabajos A Futuro

La criptografía en curvas elípticas es un terreno fértil de estudio, algunos trabajos que se identificaron en este estudio son:

– Aplicaciones de curvas elípticas en blockchain y criptomonedas.

– Desafíos y futuro de la computación cuántica en relación con curvas elípticas.

References

[1] Thaku, P and Rana, Anurag, *A Symmetrical Key Cryptography Analysis using Blowfish Algorithm.* International Journal of Engineering Research & Technology (IJERT), ISSN, 2016.
[2] Koblitz, Neal, *A course in number theory and cryptography*, Springer Science & Business Media,vol. 117, 1994.
[3] Diffie, Whitfield, *New direction in cryptography*, IEEE Trans. Inform. Theory,vol. 22, 1976.
[4] Johnson, Don and Menezes, Alfred and Vanstone, Scott, *The elliptic curve digital signature algorithm (ECDSA)*, International journal of information security, Springer, 2001.
[5] Hankerson, Darrel and Menezes, Alfred J and Vanstone, Scott,*Guide to elliptic curve cryptography*,Springer Science & Business Media, 2006.
[6] Rivest, Ronald L and Shamir, Adi and Adleman, Leonard, *A method for obtaining digital signatures and public-key cryptosystems*, Communications of the ACM, vol 21, ACM New York, NY, USA, 1978.
[7] Al Hasib, Abdullah and Haque, Abul Ahsan Md Mahmudul, *A comparative study of the performance and security issues of AES and RSA cryptography*, vol 2, IEEE, 2008.
[8] Toradmalle, Dhanashree and Singh, Rohan and Shastri, Het and Naik, Nikita and Panchidi, Vishal, *Prominence of ECDSA over RSA digital signature algorithm*, 2018 2nd International Conference on I-SMAC, IEEE, 2018.
[9] McCurley, Kevin S, *The discrete logarithm problem*, 1990.
[10] Koblitz, Neal, *Elliptic curve cryptosystems*,Mathematics of computation, vol 48, 1987.
[11] Miller, Victor S, *Use of elliptic curves in cryptography*,Conference on the theory and application of cryptographic techniques,Springer, 1978.
[12] Minfeng, Fu and Wei, Chen, *Elliptic curve cryptosystem ElGamal encryption and transmission scheme*,2010 International Conference on Computer Application and System Modeling (ICCASM 2010), vol 6, IEEE, 2010.
[13] Blake, Ian, Gadiel Seroussi, and Nigel Smart. *Elliptic curves in cryptography*. Vol. 265. Cambridge university press, 1999.
[14] Jacobson Jr, Michael John, and Prabhat Kushwaha. "Removable weak keys for discrete logarithm-based cryptography." Journal of Cryptographic Engineering 11, no. 2 (2021): 181-195.

Dirección actual: Departamento de Ciencias de la Universidad Nacional Autónoma de Honduras
Dirección de correo electrónico: `hectorflores@unah.hn`

MODELOS LINEALES DINÁMICOS APLICADO AL INDICE DE PRECIOS AL CONSUMIDOR EN HONDURAS

PEDRO JOSÉ MOLINA MORALES

Resumen. En el presente artículo se presenta el estudio de modelos lineales dinámicos aplicado al índice de precios al consumidor (IPC) en Honduras, donde primeramente se describe el modelo como uno de los casos mas importantes de un modelo en espacio de estado por su particularidad de especificarse por una distribución a priori normal, además de ciertos tipos de modelos lineales dinámicos que logran modelar diferentes tipos de series de tiempo. Se visualizó que el modelo ajustó muy bien la serie de tiempo aplicando filtros de Kalman donde nos permite calcular los estados continuamente, también se estimó de manera óptima los valores del pasado (suavizamiento) por los filtros de Kalman, y a su vez se obtuvieron pronósticos acertados, dejando en evidencia que estos modelos pueden aplicarse a cualquier tipo de serie de tiempo, dándonos resultados deseados, gracias a su flexibilidad de captar con facilidad las tendencias y estacionalidades.

Palabras Clave: Modelos lineales dinámicos, filtración, suavizamiento, pronósticos

1. Introducción

Este trabajo presenta la modelación de series de tiempo aplicado en el indicador de precio de consumidor en Honduras mediante modelos lineales dinámicos (DLM), ya que este modelo nos permite con mayor flexibilidad el modelaje de series de tiempo no estacionarias.

La importancia de modelar series de tiempo es que a través de datos históricos en el tiempo, se puedan realizar predicciones del futuro de manera precisa, por ende la aplicación del modelo mencionado anteriormente. Para ello, como toda modelación de un estudio de interés, comúnmente nos encontramos con parámetros desconocidos, lo cual se pueden estimar de diversas formas, donde en nuestro caso utilizaremos un enfoque Bayesiano [1].

Una solución a los problemas de predicción la dio Kalman (1960), en la cual utilizó una representación Bode-Shannon de procesos aleatorios y estados de transición, lo que es un método de análisis de sistemas dinámicos, lo cual fue aplicado a estos modelos lineales dinámicos. La solución fue conocida como filtro de Kalman, donde el filtrado se considera que los estados se pueden ir actualizando continuamente para conocer el estado futuro [6]. Los problemas de predicción también han sido resueltos con método de Monte Carlo con enfoque bayesiano [6].

Fecha: Agosto, 2023.

El modelo lineal dinámico (DLM), es un caso particular de los modelos de espacio de estado, donde está compuesto de dos ecuaciones, la ecuación de observación (define la distribución muestral de las observaciones condicionado a los parámetros) y la ecuación de estados (define la evolución o estado de los parámetros). Además, que se considera la perturbación de errores aleatorios, lo cual nos lleva a una interpretación probabilística del modelo [1].

Los modelos lineales dinámicos, han tenido diversas aplicaciones en marketing [3] como ser predicción de ventas, la publicidad televisiva y su impacto en la percepción del consumidor, modelaje basado en fundaciones microeconómicas, desarrollo óptimo del portafolio, entre otros. Además en [9], podemos visualizar una aplicación de un modelo lineal dinámico de enfermedades crónicas del asma, donde se desea describir patrones estacionales, concluyendo la gran utilidad de los modelos lineales dinámicos para modelar y pronosticar.

En este artículo, se aplicará el DLM al índice de precios al consumidor, donde mide los cambios de los precios y servicios que consumen los hogares, de los cuales, dichos cambios de precios afectan el poder adquisitivo real de los ingresos de los consumidores, debido a que no todos los precios cambian en la misma proporción, lo cual el índice muestra una variación promedio [7]. Los índices de precios son utilizados para comparar diferencias de niveles de precios entre distintas regiones. El IPC se calculan como promedios ponderados de las variaciones de los precios de un conjunto específico, lo cual se puede decir que tiene como objetivo medir la inflación, o bien medir el costo de vida.

2. Justificación

2.1. Lineas de Investigación. La presente investigación, dentro de las líneas de investigación prioritarias de la UNAH entra en el eje temático de ciencia, debido a que es un modelo matemático aplicado a un indicador económico como es el IPC. Además, dentro de las líneas de investigación de la maestría en Matemática, se puede encasillar en la probabilidad y procesos estocásticos, dado que se trabaja con un modelo con características Markovianas (cadenas de Markov) a una serie de tiempo en general, y que a su vez, a través de probabilidades se realizan ajustes, análisis retrospectivos y pronósticos a series de tiempo.

2.2. Importancia. El modelo lineal dinámico en un enfoque bayesiano presentado en el presente trabajo, nos ayuda a predecir y completar información a través de datos históricos, de manera precisa eventos futuros, para lograr una buena toma de decisiones, conocer valores en el pasado para entender los comportamientos y comprender el modelo.

La aplicación de estos modelos en indicadores económicos es de mucha importancia, dado que estos indicadores son variantes en el tiempo, de los cuales en su mayoría tienen tendencia y estacionalidad, lo cual los modelos de espacio de estados nos permiten captar todas estas características para comprender muy bien sus comportamientos e inferir en los parámetros de nuestro interés.

3. Antecedentes

Los modelos lineales dinámicos con un enfoque bayesiano han sido introducidos por West y Harrison (1997), donde de manera general lo define como

Definition 3.1. El modelo lineal dinámico (DLM) general está caracterízado por el conjunto de cuadruplas

$$\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$$

donde para cada t

- $F_t(n \times r)$ es una matriz conocida
- $G_t(n \times r)$ es una matriz conocida
- $V_t(n \times r)$ es una matriz de covarianza conocida
- $W_t(n \times r)$ es una matriz de covarianza conocida

La cuadrupla $\{F, G, V, W\}_t$ define la relación de Y_t al vector parámetro $\theta_t(n \times 1)$ en el tiempo t, donde durante el tiempo es secuencialmente distribuido por

$$(Y_t|\theta_t) \sim N(F_t'\theta_t, V_t),\ (\theta_t|\theta_{t-1}) \sim N(G_t\theta_{t-1}, W_t) \tag{3.1}$$

La ecuación anterior, implicítamente ya está condicionado en el conjunto de información a priori $D_{t-1} = \{D_0, Y_1, Y_2, ..., Y_{t-1}\}$, donde D_0 es el conjunto de información inicial.

Con lo visto anteriormente, de manera general se define como

$$Y_t = F_t'\theta_t + v_t, \quad v_t \sim N(0, V_t) \tag{3.2}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad \sim N(0, W_t) \tag{3.3}$$

$$(\theta_0|D_0) \sim N(m_0, C_0) \tag{3.4}$$

v_t, w_t son independientes entre si y se les conoce como los errores de la ecuación. La ecuación (3.2) se le conoce como la **ecuación de observación** y a la ecuación (3.3) como la **ecuación de estado**, y la ecuación (3.4) es la información inicial para algunos valores a prioris m_0 y C_0.

El interés en el análisis de series de tiempo, también es conocer detalladamente lo que sucedió anteriormente, a este análisis se le llama retrospectivo, lo cual nos ayuda a comprender mejor, desarrollar el modelo y mejorarlo.

Al uso de la data reciente para inferir en los estados actuales con respecto a la información actual, es llamado filtración. Entonces la distribución de interés es $(\theta_t|D_t)$, lo cual se utilizan las ecuaciones de actualización.

Un enfoque mas detallado sobre los filtrajes en particular el filtro de Kalman lo da Thomas Moore (1979), en la cual detalla ciertas caracteríticas de este filtraje para la solución del modelo, donde el problema a resolver es estimar

$$\hat{\theta}_{t|t-1} = E[\theta_t|Y_{t-1}] \quad \hat{\theta}_{t|t} = E[\theta_t|Y_t] \tag{3.5}$$

mediante las ecuaciones de actualización.

Los DLM tienen una variedad de aplicaciones en diferentes campos, donde las variables a estudiar sean series de tiempo, es decir, variantes en el tiempo. En [3], se visualizan algunas de las aplicaciones en marketing para diferentes tipos de modelos lineales dinámicos como ser **Modelo Polinómico de Primer orden**, **Modelo polinómico de Segundo Orden**, **Modelo estacional**, **Modelo de Regresión Dinámico**, entre otros, donde se visualiza que tan efectiva es la predicción a través de estos modelos a eventos futuros, y a la vez, lo bien que ajustan a diferentes patrones de datos.

En [4], Carter, Kohn y Carlin utilizaron una técnica de Monte Carlo como el Gibbs Sampling para el filtro de Kalman para la resolución de un modelo de espacio de estados aplicados, lo cual en particular pueden ser aplicados a un modelo lineal dinámico.

Mientras que en [6], nos da una mayor ampliación de los modelos lineales dinámicos aplicados a series de tiempo en general, a la vez, su implementación en el paquete estadístico R.

4. Modelo en Espacio de Estado

Sean Y_t el vector de observaciones, θ_t el vector de estados, $\{e_t,\ t \geq 1\}$ y $\{u_t,\ t \geq 1\}$ errores secuenciales donde se asume que sean mezclas normales. Entonces se define la ecuación de estado lineal como las ecuaciones (3.2) y (3.3), donde F_t y G_t depende de los valores de los parámetros de la serie de tiempo y de los errores e_t y u_t.

De la ecuación anterior [6], podemos observar que el modelo completamente depende por la distribución inicial o distribución a priori, $\pi(\theta_0)$, y las densidades condicionales $\pi(\theta_t|\theta_{t-1})$ y $\pi(y_t|\theta_t)$, esto es para todo $t \geq 1$, donde

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j)$$

Se visualiza también a través de la figura 1, la independencia condicional de las variables aleatorias del modelo, que Y_t es una variable independiente en el tiempo, pero que depende del estado actual del modelo, mientras que los estados dependen del estado anterior.

5. Modelos Dinámicos Lineales

Los modelos dinámicos lineales (DLM sus siglas en inglés) son uno de los tipos y mas importantes de los modelos de espacio de estado, donde tienen la particularidad

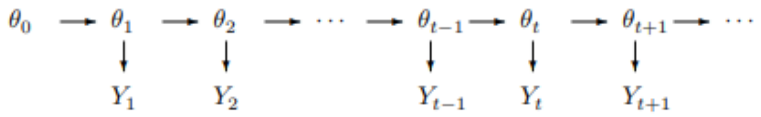


FIGURA 1. Independencia del modelo de espacio de estado

que quedan especificado por una distribución a priori normal, para el estado del vector al tiempo $t = 0$,

$$\theta_0 \sim N(m_0, C_0) \tag{5.1}$$

por ende a estos modelos también se les conoce como Modelos de Espacio de Estados Gaussianos Lineales.

Además, sean las ecuaciones (3.1) y (3.2) donde Y_t es el vector de observación, θ_t es el vector de estado, G_t y F_t son matrices conocidas, y v_t y ω_t son sucesiones de variables aleatorias independientes con distribución normal. Las ecuaciones (5.1), (5.2) y (5.3) forman el modelo dinámico lineal, donde a la ecuación (5.2) se le conoce como la ecuación de observación y a la ecuación (5.3) se le conoce como la ecuación de estados.

Como todo tipo de modelos, existen diferentes tipos de modelos que describiremos a continuación.

5.1. Modelo Polinomial de Primer Orden.

Los modelos polinomiales ajustan series de tiempo con tendencia polinómica, como ser el de primer orden que ajusta tendencias constantes. Este modelo corresponde a la especificación $\{1, 1, V, W\}$, donde $y_t \sim N(\theta_t, V_t)$ y

$$\theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t) \tag{5.2}$$

Algunas particularidades son: (i) Este es el modelo lineal dinámico mas simple y corresponde a un modelo de media variante en el tiempo; (ii) se necesitan conocer los valores V_t y W_t, donde en la práctica la mayoría de veces no es posible.

5.2. Modelo Polinomial de Segundo Orden.

El modelo polinomial de segundo orden, se obtiene con la definición de $F = (1, 0)'$ y $G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Estos modelos ajustan un crecimiento o decrecimiento lineal, es decir una tendencia lineal. En general, los modelos de k-ésimo orden tienen una tendencia de orden $k - 1$.

5.3. Modelo Estacional.

Los modelos estacionales se especifican con términos estacionales. Para patrones en diferentes tipos de periodo p, utilizamos $F = E_p = (1, 0, ..., 0)'$, con evolución definida por la matriz de permutación $G = P = \begin{pmatrix} 0 & I_{p-1} \\ 1 & 0 \end{pmatrix}$.

Alternativamente, uno puede utilizar harmónicos, para especificar patrones estacionales. El k-ésimo componente harmónico está dado por $F = E_2$, $G(k,\omega) = \begin{pmatrix} cos(k\omega) & sin(k\omega) \\ -sin(k\omega) & cos(k\omega) \end{pmatrix}$, $\omega = 2\pi/p$, $k = 1, 2, ..., h$, $h = p/2$ en caso que p sea par y $h = (p-1)/2$ si p es impar.

5.4. Modelos de regresión dinámico.

Suponga el par de valores (y_t, x_t), $t = 1, ..., T$ es observada. $\{F, I_2, V, W\}$ es el modelo de regresión dinámico, con $F_t = (1, x_t)$. El modelo de regresión dinámica se aproxima a una verdadera relación no lineal entre x_t e y_t a nivel local. Este enfoque reconoce explícitamente que el impacto de un regresor en la variable de respuesta y_t puede cambiar con el tiempo.

6. Filtración

Suponiendo que tenemos construido el modelo, uno de los objetivos es la inferencia en el modelo, en este caso de los estados, utilizando la información disponible. A partir de ahora se asume que se tiene información hasta el tiempo t $y_{1:t}$, y se desea realizar inferencia del vector de estados, entonces necesitarmos obtener la densidad condicional $\pi(\theta_s, y_{1:t})$, para ello necesitamos *filtración*, cuando $s = t$; *pronóstico*, cuando $s > t$; y *suavizamiento (análisis retrospectivo)*, cuando $s < t$.

Para la filtración, se considera que la información se puede ir actualizando continuamente en el tiempo a través de las ecuaciones de actualización que se definen en la proposición 6.1, obteniendo la densidad condicional $\pi(\theta_t|y_{1:t})$, lo que permite resolver esto se le conoce como el filtro de Kalman.

La idea general para resolver el problema de filtrado de una manera general es

- Se obtiene la distribución predictiva del estado θ_t, dadas las observaciones $y_{1:t-1}$, usando la densidad filtrada $\pi(\theta_{t-1}, |y_{1:t-1})$ y la distribución condicional $\pi(\theta_t|\theta_{t-1})$.
- Se obtiene la distribución predictiva de la siguiente observación y_t dadas las observaciones anteriores.
- Se obtiene la distribución filtrada $\pi(\theta_t|y_{1:t})$ usando la regla de Bayes con $\pi(\theta_t|y_{1:t-1})$ como la distribución a priori y $\pi(y_t|\theta_t)$ como la verosimilitud.

Ahora bien, en caso para los DLM en específicos, el problema de filtración se resuelve a través del filtro de Kalman, que se enuncia a continuación:

Proposición 6.1 (Filtro de Kalman): Consideremos el DLM enunciado en la sección 5, y supongamos que

$$\theta_{t-1}|y_{1:t-1} \sim N(m_{t-1}, C_{t-1})$$

Entonces los siguientes resultados se cumplen

I. La distribución predictiva (del primer paso) de θ_t dado $y_{1:t-1}$ es Gaussiana con parámetros

$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1}, \quad R_t = Var(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t \tag{6.1}$$

II. La distribución predictiva (del primer paso) de Y_t dado $y_{1:t-1}$ es Gaussiana con parámetros

$$f_t = E(Y_t|y_{1:t-1}) = F_t a_t, \quad Q_t = Var(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t \tag{6.2}$$

III. La distribución filtrada θ_t de Y_t dado $y_{1:t}$ es Gaussiana con parámetros

$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t, \quad C_t = Var(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t \tag{6.3}$$

donde $e_t = Y_t - f_t$

A las ecuaciones (6.1), (6.2) y (6.3) se les conoce como las *ecuaciones de actualización*.

7. Suavizamiento (Análisis Retrospectivo)

Algunos de los problemas frecuentes en las series de tiempo es reconstruir el comportamiento del sistema a partir de los datos disponibles. Para ello se utiliza el algoritmo recursivo regresivo, con el que se obtiene la distribución condicional $\pi(\theta_t|y_{1:T})$, para cualquier $t < T$, empezando con la distribución filtrada $\pi(\theta_T, y_{1:T})$ y estimando de forma retrospectiva todos los estados. Se enuncia el algoritmo a continuación:

Proposición 7.1 (Suavizamiento de Kalman): Consideremos el DLM enunciado en la sección 5, si

$\theta_{t+1}|y_{1:T} \sim N(s_{t+1}, S_{t+1})$, entonces $\theta_t|y_{1:T} \sim N(s_t, S_t)$, donde

$$s_t = m_t + C_t G_{t+1}' R_{t+1}^{-1}(s_{t+1} - a_{t+1})$$
$$S_t = C_t - C_t G_{t+1}' R_{t+1}^{-1}(R_{t+1} - S_{t+1}) R_{t+1}' G_{t+1} C_t$$

Observemos que para poder realizar el algoritmo de suavizamiento, también hay que ir realizando el algoritmo del filtro de Kalman.

8. Pronóstico

Ahora bien, una de las características mas importantes a considerar para la utilización de un modelo aplicado a la vida real es predecir observaciones futuras para la toma de decisiones, es decir estimar Y_{t+1} a través de observaciones ya disponibles $y_{1:t}$, lo cual conlleva a encontrar la distribución $\pi(Y_{t+1}|y_{1:t})$, que se le conoce como distribución de predicción a primer paso. La idea general, es estimar el estado siguiente θ_{t+1}, y luego basados en esta información, se puede obtener Y_{t+1}, todo esto por las dependencias del modelo.

Dada la naturaleza Markoviana del modelo, la distribución filtrada en el tiempo t, sirve como la distribución inicial para la futura evolución del modelo, se presenta el siguiente algoritmo que sirve para obtener la distribución de los pronósticos de los estados y observaciones:

Proposición 8.1 Consideremos el DLM enunciado en la sección 5, y supongamos que $a_t(0) = m_t$ y $R_t(0) = C_t$, entonces los siguientes enunciados se cumplen.

I. La distribución predictiva de θ_{t+k} dado $y_{1:t-1}$ es Gaussiana con parámetros

$$a_t(k) = G_{t+k}a_t(k-1), \quad R_t(k) = G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k} \tag{8.1}$$

II. La distribución predictiva de Y_{t+k} dado $y_{1:t}$ es Gaussiano con parámetros

$$f_t(k) = F_{t+k}a_t(k), \quad Q_t(k) = F_{t+k}R_t(k)F'_{t+k} + V_{t+k} \tag{8.2}$$

Para visualizar a detalle las demostraciones de las proposiciones (6.1), (7.1) y (8.1) véase [6].

9. Resultados Numéricos

Se estimó el modelo lineal dinámico (DLM) para el Indice de Precios al Consumidor (IPC) en Honduras, el periodo muestral que se consideró fue mensual desde enero 2020 hasta abril del 2023, lo cual sería un tamaño $n = 40$. El índice de Precios al consumidor mide los cambios en los precios de los bienes y servicios que consumen los hogares, de lo cual en el cuadro 1 se muestran los estadísticos descriptivos de la serie y la figura 1 muestra el gráfico de la serie de tiempo. Observemos que el mínimo es de 337.2 (2020/ene) y que el máximo es de 412.8 (2023/abr), lo cual podemos decir que los precios han aumentado durante los 3 años y medio en un 22.42 %, lo cual nos indica que la inflación va en aumento.

Serie	**Media**	**Desv. Estándar**	**Curtosis**	**Max**	**Min**
IPC	368.33	24.85	-1.23	412.8	337.2

Cuadro 1. Resumen estadísticos del Indice de Precios al Consumidor (2020/ene-2023/abr, n=40)

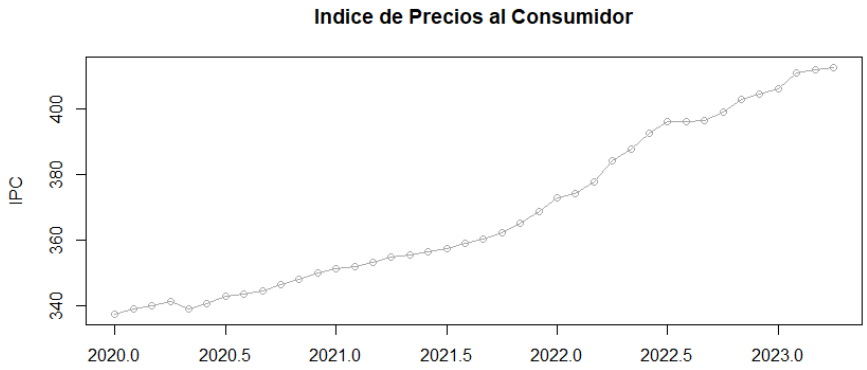


Figura 1. Indice de Precios al Consumidor (2020/ene-2023/abr, n=40)

Para la implementación del modelo, se realizó en R, utilizando la libreria *dlm* [8], donde a través de las observaciones de la serie hay que estimar los estados utilizando el filtro de Kalman, y así ajustar el modelo. Podemos visualizar que el IPC, tiene una tendencia lineal ascendente, lo cual según en la sección 5, se modelan

utilizando un modelo Polinomial de Segundo orden.

En el artículo se asume que la implemetación del modelo es con varianzas conocidas (V y W), por lo cual, $V = 2.034548e-05$ y $W = \begin{pmatrix} 1.69887 & 0 \\ 0 & 0.154142 \end{pmatrix}$ se estimaron a través del método de máxima verosimilitud. También se utilizan los valores iniciales $m_0 = (0, 0)$ y $C_0 = \begin{pmatrix} 1e07 & 0 \\ 0 & 1e07 \end{pmatrix}$.

En la figura 2 y en el cuadro 2, se puede ver que el modelo dlm de segundo orden ajusta muy bien, aplicando el filtraje visto en la sección 6.

Serie	Media	Desv. Estándar	Curtosis	Max	Min
Real	368.33	24.85	-1.23	412.8	337.2
Filtraje	368.33	24.85	-1.23	412.8	337.2

CUADRO 2. Resumen estadísticos del Indice de Precios al Consumidor Real y Filtraje (2020/ene-2023/abr, n=40)

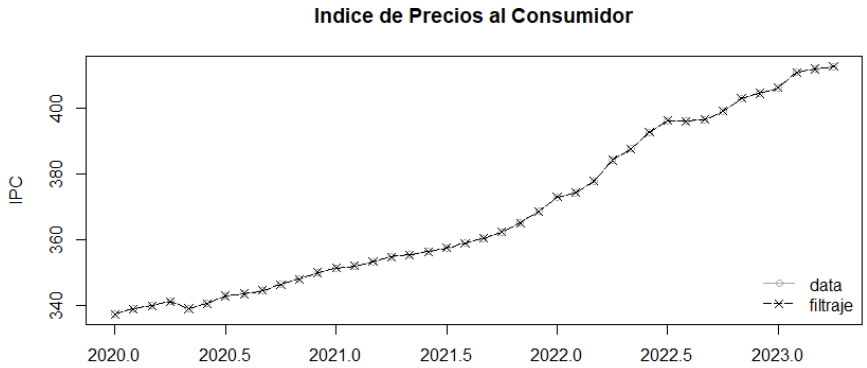


FIGURA 2. Valores Filtrados del Indice de Precios al Consumidor (2020/ene-2023/abr, n=40)

En la Figura 3, se visualiza la aplicación del suavizamiento del modelo visto en la sección 7, en la cual vemos que ajusta muy bien los datos de manera retrospectiva para entender mejor acerca del pasado del modelo.

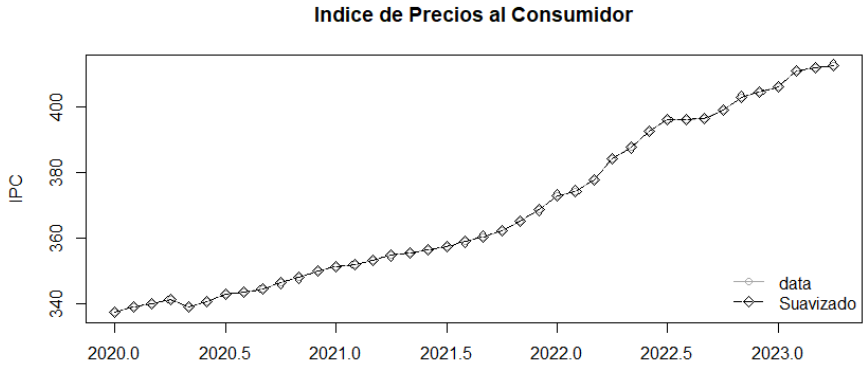


FIGURA 3. Valores suavizados del Indice de Precios al Consumidor (2020/ene-2023/abr, n=40)

Ahora bien, haremos el pronóstico para los meses mayo, junio y julio del 2023, dado que son los únicos valores a futuro que tenemos para visualizar la eficiencia de los modelos dlm utilizando las ecuaciones vistas en la sección 8. En el cuadro 3 y la figura 4, visualizamos los valores reales y pronosticados, y vemos que la variación es casi nula.

	Mayo 2023	**Junio 2023**	**Julio 2023**
Real	413.10	414.70	416.60
Pronóstico	414.77	416.75	418.72
Error	0.0040	0.0049	0.0051

CUADRO 3. Valores Pronosticados vs Valores Reales (2023/may-2023/jul)

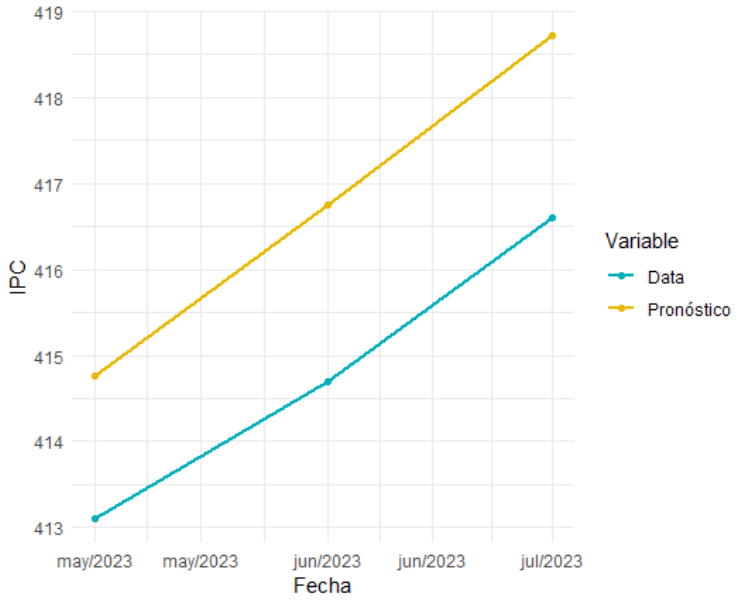


FIGURA 4. Valores Pronosticados del Indice de Precios al Consumidor (2023/may-2023/jul)

10. CONCLUSIONES

En este artículo se estimó un Modelo Lineal Dinámico (DLM) para el Índice de Precios al Consumidor (IPC) en Honduras, donde se visualiza que dichos modelos ajustan muy bien la serie de tiempo, además de realizar pronósticos bastantes acertados, descomponiendo la serie en dos ecuaciones, utilizando el método de filtraje para llevar acabo estos procesos, siempre y cuando utilicemos las varianzas V y W adecuadas, que en caso que no estén dados, una de las formas de estimarlos es con el método de máxima verosimilitud [8]. En general, con este ejemplo de aplicación podemos decir que estos modelos pueden aplicarse a todo tipo de serie de tiempo sin ninguna restricción de manera lineal y con distribución Gaussiana, permitiéndonos la obtención de pronósticos adecuados, y que a su vez, conocer su historia a través del análisis retrospectivo del modelo.

REFERENCIAS

[1] West y Harrison(1997).Bayesian Forecasting and Dynamics models.
[2] Anderson, B. Moore, J(1979). Optimal Filtering. Englewood Cliffs, New Jersey.
[3] H. Migon, M. Alves, A. Meneses,E. Pinheiro (2023). A review of Bayesian dynamic forecasting models: Applications in marketing. Universidad Federal de Rio de Janeiro.
[4] Carter, C., Kohn, R(1994).On gibbs sampling for state space models.Biometrika, 81(3), 541–553
[5] Carlin, B. Polson, N. Stoffer, D. (1992). A Monte Carlo Approach to nonnormal and nonlinear state space modeling. J. Am. Statist. Assoc.
[6] Alexandro, M, T. (2013). Modelos Dinámicos Lineales Aplicados a Series de Tiempo. Universidad Nacional Autónoma de México.

[7] Banco Mundial (2006). Manual de Precios al Consumidor.
[8] G. Petris. P. Campagnoli. S. Petrone (2009). Dynamic Linear Models with R. Springer Dordrecht Heidelberg London New York.
[9] N. Badi, M. Shakandli (2021). Application of dynamic linear model (DLM) on data corresponding to chronic asthma disease. University of Benghazi.

Departamento de Matemáticas, Universidad Nacional Autónoma de Honduras
Dirección actual: Departamento de Matemáticas, Universidad Nacional Autónoma de Honduras
Dirección de correo electrónico: pjmolina@unah.hn

ESTIMACIÓN DEL GASTO TURÍSTICO EN HONDURAS MEDIANTE MODELOS MULTINIVEL

EDUARDO S. CANALES CRUZ AND ASAEL A. MATAMOROS

Resumen. El turismo es un fenómeno social, cultural y económico que afecta a la economía de un país. En Honduras, la contribución del turismo a la economía se mide utilizando el indicador de gasto turístico, estimado por muestreo mediante un procedimiento de dos pasos. El primer paso se aplica al inicio de la visita y estima el gasto futuro del turista. Y el segundo paso se aplica al final de la visita estimando el gasto real del turista.

Hasta ahora, el segundo paso es la única herramienta utilizada para estimar el gasto del turista, y la información del primer paso se omite debido a su inexactitud natural. En este estudio, proponemos un modelo bayesiano con prioris jerárquicos, donde sus prioris se construyeron a partir de los datos del primer paso, y los posterioris se actualizaron a partir del segundo paso.

Palabras Clave: Gasto turístico, inferencia bayesiana, prioris jerárquicas, modelos multinivel

Abstract. Tourism is a social, cultural, and economic phenomenon that affects a country's economy. In Honduras, the contribution of tourism to the economy is measured using the tourism expenditure indicator, estimated by sampling using two steps procedure. The first step is applied at the beginning of the visit and estimates the tourist's future expenditure. And the second step is applied at the end of the visit estimating the actual tourist expenditure.

So far, the second step is the only tool used to estimate tourist spending, and the information from the first step is omitted due to its natural inaccuracy. In this study, we propose a Bayesian model with hierarchical priors, where its priors were constructed from the first step data, and the posterioris were updated from the second step.

Key words and phrases: Touristic expenditure, Bayesian inference, Hierarchical priors, multilevel models.

1. Introducción

El turismo es una actividad económica que en las últimas décadas se ha desarrollado de manera acelerada a nivel mundial. Según la OMT (2018) "es un sector fundamental de generación de ingresos en las economías emergentes y en desarrollo," y cada vez son más los países que dan un mayor peso al turismo en la planificación de sus políticas económicas. En Honduras, al igual que en muchos otros países se considera a la actividad turística como prioritaria e importante para dinamizar la economía mediante la atracción de inversión nacional y extranjera pero sin descuidar el concepto de desarrollo sostenible. Las estadísticas de turismo en Honduras se han centrado únicamente en el análisis a un nivel individual sin tener cuenta

Date: Agosto, 2023.

estructuras y relaciones más amplias que pueden existir al estudiar un fenómeno físico, sin embargo, muchos problemas de interés en la investigación y la práctica se desarrollan en contextos complejos donde los individuos están agrupados en grandes estructuras jerárquicas como equipos, organizaciones o regiones geográficas.

El análisis del gasto turístico es fundamental para comprender y evaluar el derrame económico que este conlleva a la economía del país. La estimación adecuada de este indicador mediante modelos estadísticos proporciona una compresión integral y precisa del gasto generado por parte de los turistas que ingresan a realizar actividades económicas en un país. El gasto turístico varia según la zona visitada debido a que se realizan diferentes actividades económicas en cada una de ellas, ver figura 2. Zonas muy populares como la zona insular del país generan alto costos por hospedaje debido a su alta demanda turística. Esto genera una gran variabilidad en el gasto dificultando el proceso de estimación, y métodos clásicos como una media muestral [16] o cualquier estimación global del gasto producen valores poco confiables e inexactos. Un fenómeno usual al estudiar estadísticas de turismo son los gastos atípicos generados por los turistas con alto poder adquisitivo, ver figura 1. Estos valores atípicos hacen que la distribución del gasto sea de colas pesadas, invalidando los supuestos de normalidad utilizados en los estimadores de medias muéstrales, y dificultando las comparaciones entre grupos, ver figura 2.

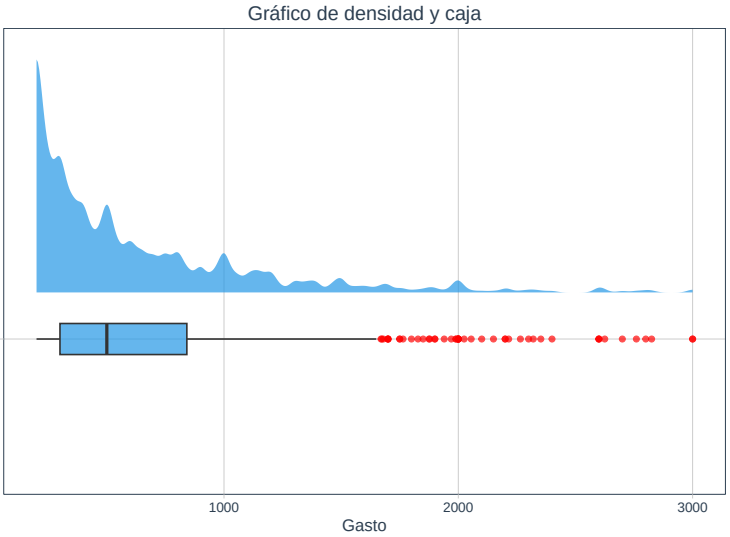


FIGURA 1. Distribución del gasto turístico, donde aproximadamente el 11 % representan valores atípicos o extremos,según los datos recolectados por la encuesta (EGPV) del [16].

En ese sentido, para obtener estimaciones precisas es necesario utilizar enfoques estadísticos adecuados que permiten considerar la estructura jerárquica presente en el estudio del gasto por estadía generado por los turistas que visitan el país, capturando la variabilidad entre los turistas como individuos y teniendo en cuenta la estructura jerárquica presente según el destino principal que estos visitan. Los modelos multinivel permiten incorporar estructuras jerárquicas en el análisis

teniendo en cuenta los efectos fijos como los efectos aleatorios lo cual los hace especialmente útiles para analizar datos longitudinales [19], datos de medidas repetidas o agrupados en diferentes niveles.

El [16] en su boletín de estadísticas presenta las tendencias del sector turismo en Honduras durante el periodo 2012-2016 el cual es de mucha importancia para este trabajo, donde se usan los datos generados por la encuesta del gasto y perfil del visitante, 2016 (EGPV). Dicha encuesta extrae el gasto por estadía generado por los turistas que ingresan al país, y se realiza al momento que el turista ha realizado una gran parte de su estadía y esta cerca de regresar a su país de residencia.

Este estudio propone un modelo probabilístico *log-normal* multinivel estratificado por la zona visitada por los turistas, con medias entre grupos y varianza desconocidas. El modelo propuesto obtiene estimaciones precisas del gasto turístico en Honduras y las colas pesadas de la distribución log-normal permite tener estimaciones resistente a gastos atípicos. Adicionalmente, la estructura multinivel del modelo permite determinar la variabilidad del gasto en los diferentes destinos turísticos del país como estructura jerárquica, con el fin de proporcionar información más precisa y relevante para la planificación estratégica para impulsar el crecimiento y desarrollo del sector turismo. La inferencia de los parámetros desconocidos se realizó mediante un enfoque Bayesiano con distribuciones a priori débiles y poco informativas [10], y las distribuciones a posteriori del modelo propuesto son muestreadas utilizando un Monte Carlo Hamiltoneano [6]; mediante el algoritmo NUTS [14] realizando las implementaciones en Stan [29], mediante el software computacional R [28]. Nuestro modelo fue comparado con otros modelos recientes de la literatura que al ser modelos globales no contemplan la variabilidad entre grupos que presentan los datos. Comparamos la capacidad predictiva de los modelos utilizando validación cruzada [32] y el criterio de información de Watanabe [34], donde nuestro modelo superó a los modelos de la literatura, siendo este el modelo con mejor capacidad predictiva.

El resto del documento se resumen a continuación. La sección 2 aborda la justificación del trabajo. En la sección 3, presenta los modelos multinivel y sus aplicaciones en diferentes áreas de investigación, y se presentan diferentes estudios del gasto turístico en diferentes países. En la sección 4 hacemos una descripción estadística del Gasto turístico en Honduras y presentamos el modelo multinivel propuesto, como el modelo reciente propuesto por [12]. En la quinta sección presentamos los resultados obtenidos al implementar nuestro modelo y el modelo de [12] al gasto registrado por los turistas en Honduras en el año 2016. Finalmente, en el sección 6 presentamos las conclusiones de las ventajas de nuestro modelo sobre los demás propuestos en la literatura y por ultimo trabajos a futuro.

2. Justificación

La Universidad Nacional Autónoma de Honduras en sus programas de maestrías presentan cuatro ejes prioritarios de investigación en los cuales se deben enfocar las investigaciones realizadas, este trabajo se encuentra en el *Eje de investigación: Desarrollo Económico y Social* y dentro de las lineas de investigación de la Maestría en Matemáticas con Orientación en Estadística Matemática: Estadística multivariada y modelos lineales generalizados.

3. Antecedentes

Los modelos multinivel, también conocidos como modelos de efectos mixtos son ampliamente utilizados para el análisis de datos, estos modelos permiten tener en cuenta los efectos fijos como los efectos aleatorios lo cual los hace especialmente útiles para analizar datos longitudinales [19] o de panel [35, 23], datos de medidas repetidas o agrupados en diferentes niveles. En las ultimas décadas las metodologías estadísticas para analizar medidas repetidas han tenido un notable desarrollo debido a la facilidad de su implementación gracias a los avances del la computación.

[17] introdujo los modelos multinivel y su aplicación en el análisis de datos longitudinales que tiene en cuenta los efectos fijos como los efectos aleatorios, este trabajo sentó bases para el desarrollo de los modelos mixtos en diversas áreas de investigación. Adicionalmente, las estimaciones mediante los modelos multinivel se pueden obtener utilizando diferentes métodos; [4] aborda cada una de estas y explica en detalle como pueden ser obtenidas a partir de medidas repetidas o longitudinales. Los primeros métodos numéricos para la estimación en los modelos multinivel son mínimos cuadrados [1]; máxima verosimilitud [18]; y máxima verosimilitud restringida [19]. En la actualidad los métodos Bayesianos se han vuelto popular para estimar modelos con estructuras multinivel ya que permiten obtener estimaciones confiables de los efectos fijos y aleatorios mediante la introducción de información adicional a través de la distribución a priori, y permiten cuantificar la incertidumbre de los efectos mediante el uso de la distribución a posteriori [3].

Estos modelos son una extensión de los modelos de regresión lineal que acoplan varios modelos lineales para cada nivel de análisis, es decir considerar dentro un mismo modelo los distintos niveles de la estructura jerárquica y conocer la variabilidad debida al segundo nivel [7], del mismo modo un modelo Bayesiano multinivel se construye a partir del modelo de regresión lineal ordinario y intentará predecir la variable de respuesta (y_i) mediante una combinación lineal de un intercepto y una pendiente que cuantifica la influencia de un predictor (x_i), para mas detalles ver [24]. Las siguientes ecuaciones muestran la estructura clásica de un modelo lineal simple,

$$y_i = \alpha + \beta x_i + \sigma_e e_i, \quad e_i \sim N(0,1). \tag{3.1}$$

En este modelo las variables de respuesta y_i se distribuyen normalmente alrededor de la media $\alpha + \beta x_i$ y varianza residual σ_e^2. El modelo anterior se puede extender al siguiente modelo multinivel con J niveles o grupos, incorporando un intercepto variable expresado a continuación:

$$y_{ij} = \alpha_j + \beta x_i + \sigma_e e_i, \quad e_i \sim N(0,1); \tag{3.2}$$

$$\alpha_j \sim N(\alpha, \sigma_\alpha^2) \quad \text{para todo } j \in 1, 2, \ldots, J. \tag{3.3}$$

donde α_j indica que a cada grupo j se le da un intercepto único, [24] menciona que además de la varianza σ_e^2 también se está estimando un componente más σ_α^2 que representa la varianza de la distribución de los interceptos variables, esta se considera como la variación del parámetro α entre los grupos j, siguiendo una metodología similar es posible introducir un termino de pendiente variables que pueda cambiar según el grupo j.

$$y_{ij} = \alpha_j + \beta_j x_{i,j} + \sigma_e e_i, \quad e_i \sim N(0,1); \tag{3.4}$$

$$\alpha_j \sim N(\alpha, \sigma_\alpha^2), \quad \beta_j \sim N(\beta, \sigma_\beta^2), \quad \text{para todo } j \in 1, 2, \ldots, J.$$

A estas pendientes variables se les asigna una distribución a priori centrada en la gran pendiente β, y con varianza σ_β^2. Con lo anterior, la inferencia Bayesiana nos permite realizar todas estas aseveraciones posibles que podemos aplicar en los modelos multinivel, para ello supongamos que tenemos K observaciones agrupadas en J grupos, la variable de interés es y_{ij} que representa la observación $i-$ésima en el grupo j, podemos modelar la variable y_{ij} en un modelo multinivel lineal Bayesiano de la siguiente manera;

$$y_{ij} = \mu_j + \beta_j x_i + \sigma_e e_i, \quad e_i \sim N(0,1); \tag{3.5}$$

$$\mu_j \sim N(\mu, \sigma_\mu^2),\ \beta_j \sim N(\beta, \sigma_\beta), \quad \sigma_e \sim \text{student-t}(v_e),$$

con distribuciones a priori:

$$\mu \sim N(\mu_0, \sigma_{\mu 0}^2),\ \beta \sim N(\beta_0, \sigma_{\beta 0}^2), \quad \sigma_\mu \sim \text{student-t}(v_0),\ \sigma_\beta \sim \text{student-t}(v_1).$$

donde, y_{ij} es la observación i-ésima en el grupo j, μ_j es la media del grupo j para la observación i centrada en la media global μ y escala σ_μ; y σ_e^2 es la varianza asociada con las observaciones individuales, también conocida como varianza residual o error relativo. Los valores $\mu_0, \beta_0, \sigma_{\mu 0}, \sigma_{\beta 0}, v_e, v_0$ y v_1 son hiper-parámetros conocidos y elegidos por el investigador.

Las actividades turísticas se consideran como una de las fuentes más importantes en el crecimiento económico de un país es por ello que la estimación del gasto turístico es un aspecto fundamental para comprender el impacto y poder tomar decisiones adecuadas. En base a ello se han realizado diversos trabajos a lo largo del tiempo, como [15] el cual es uno de los primeros trabajos en aplicar un análisis multinivel de los determinantes de gasto turístico de los hogares donde los resultados obtenidos indican que las variables como edad, renta familiar, la propiedad de un vehículo y el uso del internet influyen positivamente en el gasto turístico, [35] propone un modelo de regresión múltiple para una cuantificación del gasto turístico, asimismo [33] trabajo con determinantes del gasto turístico aplicando varios modelos multinivel.

Uno de los problemas que se generan al momento de realizar estimaciones del gasto turístico es que se obtienen colas pesadas en las distribuciones, esto debido a que hay muchos factores que pueden influir [12]. A lo largo de estos años muchos estudios se han enfocado en abordar este problema de colas pesadas; [5] mencionan que una solución es segmentar el gasto turístico por categorías en el cual identificaron asociaciones estadísticamente significativas entre los distintos segmentos de gasto, examinando la importancia de una serie de variables socio-económicas y de comportamiento. [11] proponen que otra forma es dividir el gasto total de los turistas basados en el gasto según el país de origen y destino que estos tomen, por otra parte, en estudios más recientes como [22] utilizaron series temporales para pronosticar el gasto medio de turista en España.

El modelo presentado por [12], cumple con todas las características de nuestra investigación dado que sera un punto de comparación con nuestro modelo propuesto, [12] prueban en su artículo que realizando una reparametrización de la distribución *log-skew normal* de tres parámetros para la modelización del gasto turístico en base usando distintas covariables como el país de origen, destino y el gasto total, se obtienen resultados satisfactorios en los datos del gasto en las las partes de la distribución empírica, de igual forma el modelo se adapta bien para captar la

asimetría, curtosis y colas pesadas que las tres variables mencionadas tienden a presentar en la practica, como lo es en nuestro caso, ver figura 2. En tal sentido por ello utilizamos modelos multinivel y así poder obtener estimaciones precisas en la variabilidad del gasto turístico.

4. Modelización del gasto turístico por estadía en Honduras

La encuesta de gasto y perfil del visitante (EGPV) para el año 2016 se aplicó a $4,712$ turistas extranjeros, de los cuales $2,303$ de ellos declararon un gasto válido para el análisis, alrededor del $50\,\%$ de los turistas pernoctaron al menos cuatro noches en el país y en promedio pernoctaron nueve noches (*estadísticas del número de noches por estadía; min. 1, media 9, mediana 4, max. 300*). El gasto turístico promedio es de 323 dólares, el gasto mínimo reportado fue de 50 centavos de dolar, mientras que el gasto máximo fue de $10,000$ dólares. Además, el $13\,\%$ (292 *turistas*) de los turistas encuestados reportaron un gasto mayor a 700 dólares, esto es, el doble del gasto promedio registrado para el año 2016.

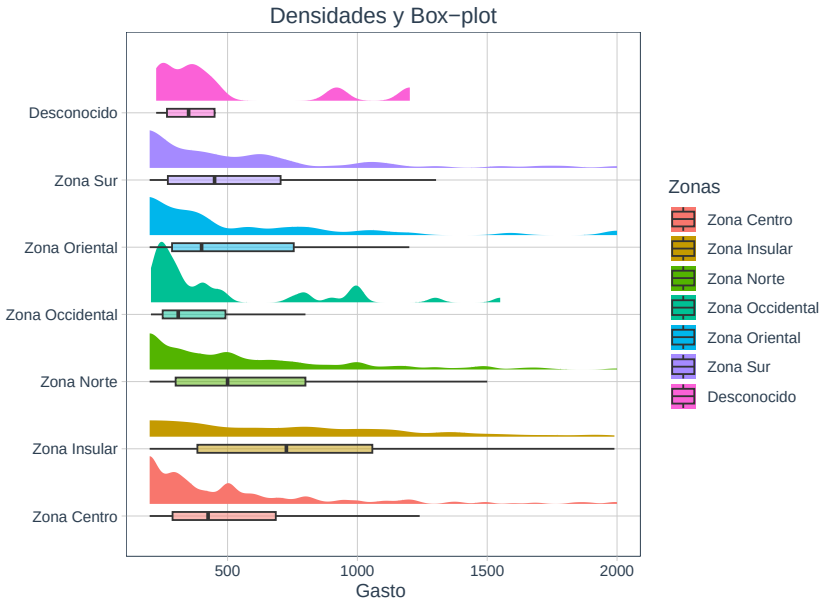


Figura 2. Distribución del gasto turístico por estadía desagregado por zonas que visitaron en Honduras, según los datos recolectados por la encuesta (EGPV) del [16].

El cuadro 1 proporciona un resumen detallado del gasto turístico de los visitantes sin aplicarle la transformación logarítmica a los datos, desglosado por las diferentes zonas que visitaron en Honduras, cada una de ella exhibe características únicas en términos de patrones de gasto, mencionar que las zonas están divididas según la (EGPV) de la siguiente manera; *Zona centro, insular, norte, occidental, oriental, sur y desconocida*. Por ejemplo, la *zona centro* muestra una concentración moderada del gasto al rededor de la mediana indicando una distribución relativamente equilibrada, caso contrario en la *zona insular* donde se observa una mediana más

alta de 733, sugiriendo una concentración de gastos excepcionalmente altos y de valores atípicos. Además, las *zonas norte, oriental* y *sur* muestran una asimetría notable indicando una distribución desplazada ó colas pesadas hacia valores superiores. La *zona desconocida*, el valor de la mediana en términos de gasto turístico se sitúa en un punto intermedio, lo que sugiere una mayor uniformidad en los gastos.

Al analizar detalladamente la asimetría en cada zona, logramos identificar patrones en la distribución del gasto turístico; comenzando con la *zona centro* donde observa una asimetría de 5.16, indicando que la distribución tiene una cola más larga hacia los valores superiores, lo mismo sucede en las *zonas insular y norte* con valores de 3.93 y 7.16 respectivamente, lo que sugiere que hay presencia de valores atípicos, la *zona oriental* también muestra una asimetría más alta de 7.65, reflejando una cola más larga hacia la derecha, en la caso de la *zona occidental y sur* indican colas pesada pero no tan pronunciadas, *zona desconocida* presenta una asimetría de 1.71, sugiriendo una distribución menos sesgada y más cercana a la simetría en comparación con otras zonas. En cuanto a la kurtósis, se notan variaciones en la forma de las distribuciones en el caso de las *zonas centro y oriental* exhiben colas pesadas y picos más notables respectivamente, lo cual implica que hay mayor concentración de valores extremos, la *zona norte* destaca con una kurtósis excepcionalmente alta de 84, las *zonas occidental y sur* presentan colas más acentuadas a la distribución normal, aunque menos extremas, lo mismo en la *zona insular*, por último la *zona desconocida* indica una distribución con colas menos pronunciadas y picos menos agudos.

CUADRO 1. Resumen del gasto turístico desagregado por zonas, de los visitantes que ingresaron al país año 2016.

Zona	n	min.	sd	media	mediana	max.	asimetría	kurtósis
Centro	701	3.75	594	333	140	7,500	5.16	40.4
Insular	150	110	872	899	733	7,500	3.93	23.1
Norte	528	0.5	673	438	235	10,000	7.16	84
Occidental	326	2	187	114	55	1,550	4.23	21.6
Oriental	210	2	546	204	50	6,300	7.65	75.4
Sur	370	1.33	363	161	50	3,000	4.77	26.4
Desconocida	18	3.25	326	271	193	1,202	1.71	2.27
Global	2303	0.5	592.17	323.2	116.7	10,000	6.07	62.74

En la figura 2, se presenta las densidades del gasto turístico desagregado por zonas, donde se observa que la mayor parte de la información del gasto se concentra entre los primeros mil valores, el gráfico de cajas revela muy poca información, más del 80 % de los datos están cercanos a cero, estos gráficos indican que el gasto turístico no sigue una distribución normal. En ese sentido, se propone aplicar una transformación logarítmica a los datos para lidiar con los valores extremos de la muestra, el cual representan aproximadamente el 11 %, ver figura 1. En la figura 3, se presentan la transformación logarítmica del gasto desagregado por zonas, donde las densidades indican que los datos siguen una distribución normal, el gráfico de cuantiles mejora significativamente al aplicar dicha transformación en comparación con la figura 2.

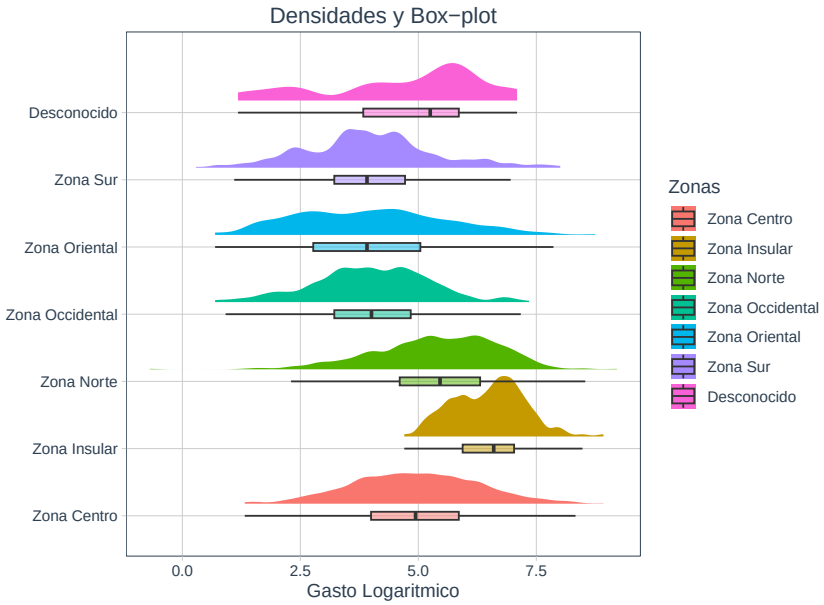


FIGURA 3. Distribución del gasto turístico logarítmico por estadía desagregado por zonas que visitaron en Honduras, según los datos recolectados por la encuesta (EGPV) del [16].

En muchas situaciones, los datos empíricos muestran una ligera o marcada asimetría y colas pesadas, que refleja valores extremos [12], es por ello, que en este trabajo proponemos un modelo jerárquico utilizando un solo nivel con los datos en escala logarítmica. Sea $y_{ij} = \{y_{1,j_1}, y_{2,j_2}, \ldots, y_{n,j_n}\}$ una muestra aleatoria para el gasto turístico, donde y_{ij} es el gasto para la observación i en la zona j, tal que $j = 1, 2, 3, \ldots, 7$; y cada observación en escala logarítmica, se distribuye normal jerárquicas con vector de medias entre grupos $\mu = (\mu_1, \mu_2, \ldots, \mu_7)$ y varianza global (σ^2) desconocidas,

$$\log(y_{ij}) \sim N(\mu_j, \sigma), \tag{4.1}$$

con distribuciones a priori:

- $\mu_j \sim N(0, 10)$ es la media entre grupos del gasto turístico para la zona j; que siguen una distribución normal con media cero y varianza diez.
- $\sigma \sim$ student-t$(5, 0, 10)$ es la varianza global del logaritmo del gasto, que sigue un distribución t de student definida en los valores positivos, con $v = 5$ grados de libertad, centrada en cero y con escala de diez.

En este estudio utilizamos distribuciones a priori débilmente informativas, y se eligen de tal forma que dichas distribuciones provean poca información y mejoren la geometría de la función propuesta tal que el muestreo de la distribución a posteriori sea mas estable, para mas detalles ver [10]. Asimismo, realizaremos una comparación con la distribución propuesta por [12], el cual utilizó un modelo *log-normal asimétrico* para estimar el gasto turístico global. El modelo viene dado de

la siguiente manera:

$$\log(y_i) \sim \text{skew-N}(\mu, \sigma, \alpha), \tag{4.2}$$

con parámetros de locación (μ), escala (σ) y de forma (α) desconocidos, y distribuciones a priori:

- $\mu \sim N(0, 10)$ es la media del gasto turístico global; que siguen una distribución normal con media cero y varianza diez.
- $\sigma \sim$ student-t$(3, 0, 10)$ es la varianza global del logaritmo del gasto, que sigue un distribución t de student definida en los valores positivos, con $v = 3$ grados de libertad, centrada en cero y con escala de diez.
- $\alpha \sim N(0, 1)$ es la asimetría global del logaritmo del gasto, que sigue un distribución normal estándar.

Donde $i = 1, \ldots, n$ son los índices de los datos globales, n denota el número de observaciones, skew-N() denota la densidad de una distribución normal asimétrica, y las distribuciones a priori elegidas son débilmente informativas.

Para realizar las estimaciones del gasto turístico de los modelos propuestos, utilizaremos métodos de Markov Chain Monte Carlo (MCMC) como [21, 13]. En particular emplearemos un Monte Carlo Hamiltoneano [6, 14] implementado en el lenguaje de programación probabilista [26], y ejecutado en el lenguaje de programación R [27, 2]. Para cada implementación se corrieron cuatro cadenas con valores iniciales distintas, con un total de 2,000 iteraciones por cadena, eliminando las primeras 1,000 iteraciones por cadena (*warm-up*). Presentamos las distribuciones a posteriori del modelo multinivel propuesto y evaluamos la convergencia de las cadenas utilizando el factor de convergencia ($\hat{R}$), y los tamaños de muestra efectivos (ess) [31]. Para las visualizaciones de las cadenas se usó trazas y gráficos de densidades [9, 8]. Para, evaluar el ajuste del modelo utilizaremos evaluación de la densidad predictiva (*posterior predictive checks*) visualizando el ajuste con gráficos de densidades, [25]. Finalmente, comparamos ambos modelos utilizando validación cruzada, ver [30]. Dicho método computa las esperanzas de las log-predictivas (*elpd*) removiendo una observación y_i de la muestra y computar la log-predictiva del modelo para dicha observación faltante; este procedimiento es caro computacionalmente pero puede ser aproximado usando un muestreo por importancia con un suavizado de Pareto [32], o por submuestreo [20].

5. Análisis y Resultados

En el cuadro 2 se proporciona un resumen de los resultados obtenidos de las distribuciones a posteriori del modelo *log normal* multinivel. En primer lugar la media representa el estimador puntual de los parámetros al igual que la mediana. Los valores de la desviación (sd) y desviación absoluta media (mad) son los estimadores del error de Monte Carlo, valores bien cercanos a cero reflejan que las estimaciones obtenidas tienen un alto grado de confiabilidad y se aprecia que son estables en todas las zonas analizadas, los valores *Q5* y *Q95* señalan los intervalos de credibilidad al 10 %, el indicador *Rhat* se refiere al diagnóstico de convergencia para comparar las estimaciones entre y dentro de la cadenas de cada parámetros estimado, valores muy cercanos a uno implican un buen diagnostico, y lo interpretamos como que las cadenas del MCMC muestrearon y convergieron eficientemente.

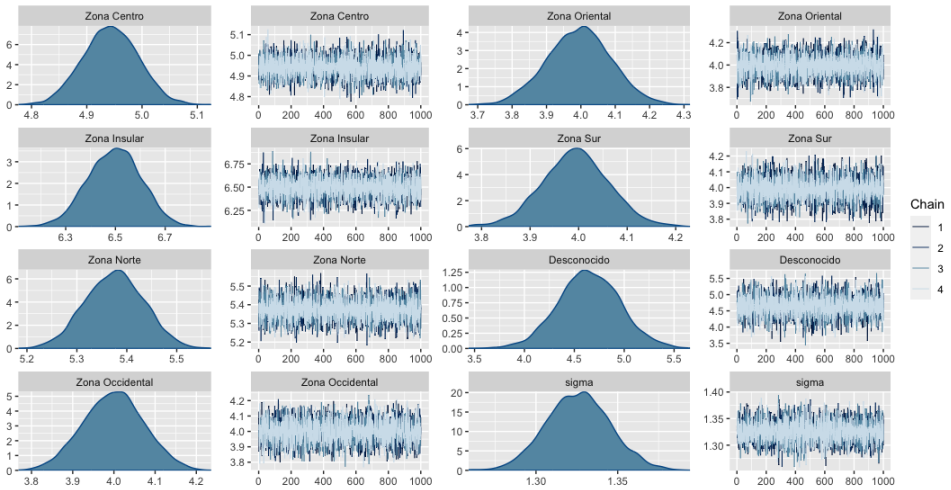


FIGURA 4. Densidades y cadenas de Markov del modelo *log normal* multinivel propuesto

CUADRO 2. Resumen de las distribuciones a posteriori obtenidas del modelo *log normal* multinivel, las zonas establecidas en la columna "variable" representan las medias grupales del modelo multinivel que corresponden a las zonas establecidas en los datos.

variable	media	mediana	sd	mad	q5	q95	Rhat	ess_bulk	ess_tail
Zona Centro	4.94	4.94	0.05	0.05	4.86	5.02	1.00	6,895	3,349
Zona Insular	6.50	6.50	0.11	0.11	6.32	6.67	1.00	8,271	3,389
Zona Norte	5.38	5.38	0.06	0.06	5.28	5.47	1.00	6,516	2,820
Zona Occidental	4.00	4.00	0.07	0.07	3.88	4.12	1.00	6,931	3,000
Zona Oriental	4.00	4.00	0.09	0.09	3.84	4.15	1.00	7,455	2,749
Zona Sur	3.99	3.99	0.07	0.07	3.88	4.11	1.00	6,738	3,052
Desconocido	4.65	4.65	0.31	0.31	4.14	5.17	1.00	7,049	3,090
σ	1.33	1.33	0.02	0.02	1.29	1.36	1.00	7,784	2,982

Los valores *ess_bulk* y *ess_tail*, son indicadores para medir la eficiencia de las estimaciones de tamaño de muestra efectivo, la primera se enfoca en la regiones donde la densidad de probabilidad es más alta es decir la parte central de la distribución y la segunda en las colas o extremos de la distribución, estos valores deben ser cercanos a 4000 o similares entre si, en nuestro caso los resultados indican que en cada zona refleja la eficiencia con lo que las cadenas de están explorando y muestreando en ambas partes de la distribución. Finalmente, los valores obtenidos se observan bien y los indicadores de convergencia muestrean bien nuestro modelo propuesto para la estimación del gasto turístico el cual se puede observar en la figura 4, donde las cadenas MCMC lograron hacer un *"mixing"* efectivo al juntarse entre sí y parecer estacionaras, las densidades se ven simétricas y sin múltiples modas, siendo estos indicadores de convergencia.

Los resultados que se obtuvieron del modelo de [12], todos los *Rhat* son cercanos a 1 siguiendo la misma explicación de nuestro modelo, asimismo los *ess_bulk* y *ess_tail* también son cercanos a las 4000 iteraciones obtenidas, por lo tanto, aceptamos la convergencia del modelo.

CUADRO 3. Resumen de la comparación del modelo log normal multinivel y [12]

modelo	elpd_diff	se_diff	elpd_loo	waic
multi-nivel	0	0	-3904.12	7808.23
Gomez	-294	22.29	-4198.12	8396.24

El cuadro 3 muestra los resultados obtenidos por validación cruzada, la primer columna representa la diferencia de las log predictivas esperadas *elpd_diff* como se observa se la diferencia es de 294 unidades a favor del modelo multinivel por lo cual nuestro modelo propuesto tiene mayor capacidad predictiva, la segunda columna *sd_diff* corresponde al error estándar de las diferencias de estimación del valor *elpd_diff*, en este caso la diferencia es poca, por lo tanto se acepta el ajuste proporcionado. La tercera columna *elpd_loo* son los valores de la log predicitivas de cada modelo, mediante el método de validación cruzada *LOO ((Leave-One-Out)* el cual indica la calidad predictiva, se observa que nuestro modelo presenta mejor resultados. Por último tenemos el valor del criterio de información de Watanabe-Akaike *WAIC* [34], dicho criterio elige al modelo con valor menor, el cual confirma que nuestro modelo es el que presenta mejores resultados.

6. CONCLUSIONES

En la comparación entre el modelo multinivel propuesto y el modelo de referencia presentado por [12] se observa una serie de diferencias sustanciales que respaldan la viabilidad y ventajas de nuestro enfoque. Una característica destacada del modelo propuesto es la eliminación de la necesidad de estimar un parámetro de forma, el cual suele ser bien engorroso de estimar, esto simplifica el proceso de estimación y refuerza la robustez de las predicciones; en cambio el modelo de [12] asume una uniformidad del gasto en todo el país sin considerar las variaciones que se presenten en las distintas zonas turísticas es decir para los datos del gasto turístico global como se observa en la figura 1. Nuestro modelo multinivel si aborda explícitamente estas diferencias que surgen con lo cual se obtienen estimaciones más precisas y ajustadas como se observó en los resultados de la validación cruzada. Asimismo, es importante resaltar que según el *LOO ((Leave-One-Out)* obtenido nuestro modelo puede predecir una nueva información con precisión.

En este trabajo, hemos profundizado la relevancia que tienen las distribuciones *log normales* y los modelos multinivel. Las distribuciones *log normales* son especialmente útiles para modelar datos en los cuales se cuenta con colas pesadas como se observó en la figura 2 del gasto turístico, permitiendo capturar de manera efectiva los valores atípicos o extremos y al aplicar transformaciones logarítmicas se logró regularizar las distribuciones mejorando la estabilidad y precisión de las estimaciones. En relación a los modelos multinivel hemos podido estudiar la capacidad que estos tienen para identificar la variabilidad y estructura jerárquica de los datos,

proporcionando una comprensión enriquecedora y contextualizada de los factores que pueden influir en diversas situaciones de estudio.

7. Trabajo a Futuro

Para mejorar el modelo *log-normal* multinivel utilizado en este trabajo, se incorporarán covariables al modelo propuesto, donde se podrá obtener una análisis más completo, como por ejemplo el país de residencia de los turistas el cual podría influir en los patrones del gasto turístico desagregado por zonas. Asimismo, considerar otras distribuciones para el análisis de la pesadez en las colas del gasto turístico, con ello nos permitirán capturar y entender mejor el comportamiento de los datos, una de ellas es la distribución t de student.

Referencias

[1] Douglas M Bates and Donald G Watts. *Nonlinear regression analysis and its applications*. Wiley, 1988.

[2] Paul-Christian Bürkner. brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28, 2017.

[3] Paul-Christian Bürkner. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1):395–411, 2018.

[4] Juan Carlos Correa Morales and Juan Carlos Salazar Uribe. Introducción a los modelos mixtos. *Escuela de estadistica*, 2016.

[5] Ruth Craggs and Peter Schofield. Expenditure-based segmentation and visitor profiling at the quays in salford, uk. *Tourism Economics*, 15(1):243–260, 2009.

[6] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

[7] Virginia Pérez Fernández. *Los modelos multinivel en el análisis de factores de riesgo de sibilancias recurrentes en lactantes: enfoques frecuentista y bayesiano*. PhD thesis, Universidad de Murcia, 2012.

[8] Jonah Gabry and Tristan Mahr. bayesplot: Plotting for bayesian models, 2019. R package version 1.7.1.

[9] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *J. R. Stat. Soc. A*, 182:389–402, 2019.

[10] Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. Bayesian workflow. *arXiv preprint arXiv:2011.01808*, 2020.

[11] Emilio Gómez-Déniz, Jorge V Pérez-Rodríguez, and José Boza-Chirino. Modelling tourist expenditure at origin and destination. *Tourism Economics*, 26(3):437–460, 2020.

[12] E. Gómez–Déniz, N. Dávila-Cárdenes, and J. Boza-Chirino. Modelling expenditure in tourism using the log-skew normal distribution. *Current Issues in Tourism*, 25(14):2357–2376, 2022.

[13] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[14] Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

[15] Wei-Ting Hung, Jui-Kuo Shang, and Fei-Ching Wang. A multilevel analysis on the determinants of household tourism expenditure. *Current Issues in Tourism*, 16(6):612–617, 2013.

[16] República de Honduras Instituto Hondureño de Turismo. Boletín de estadísticas de turismo 2012 - 2016. 2016.

[17] Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[18] Mary J. Lindstrom and Douglas M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022, 1988.

[19] Mary J. Lindstrom and Douglas M. Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673–687, 1990.

[20] Måns Magnusson, Michael Riis Andersen, Johan Jonasson, and Aki Vehtari. Leave-one-out cross-validation for bayesian model comparison in large data, 2020.

[21] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
[22] Evelyn Yanela Nolasco Palomino. Pronóstico del gasto medio por turista en España mediante el uso de series temporales. 2022.
[23] Sara A Proença and Elias Soukiazis. Demand for tourism in Portugal: A panel data approach. 2005.
[24] Lesky Ibeth Rivas Martínez and Cristian Andrés Cruz Torres. Análisis multinivel de factores que afectan el rendimiento escolar en español tercer grado en Honduras. *Paradigma: Revista de Investigación Educativa*, 29(48):93–119, dic. 2022.
[25] Teemu Säilynoja, Paul-Christian Bürkner, and Aki Vehtari. Graphical test for discrete uniformity and its applications in goodness of fit evaluation and multiple sample comparison. *arXiv preprint arXiv:2103.10522*, 2021.
[26] Development. Team Stan. Stan: Stan modeling language: User's guide and reference manual. 2017.
[27] Stan, Development. Team. Stan: A c++ library for probability and sampling, version 2.16.0. 2017.
[28] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
[29] Stan Development Team. Stan: A c++ library for probability and sampling, version 2.14. 0, 2017a.
[30] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432, Aug 2016.
[31] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc, 2020. Advance publication.
[32] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling, 2015.
[33] Julio Vena Oya. Determinantes del gasto efectivo en el turismo cultural. 2020.
[34] Sumio Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 2010.
[35] Leire Zubizarreta Barrenetxea. Análisis estadístico del gasto medio del turista en España y los factores que lo determinan: Una propuesta de análisis de datos de panel. 2020.

Department of Mathematics, Universidad Nacional Autónoma de Honduras

Email address: escanales@unah.hn

Computer Science Department, Aalto University

Email address: izhar.alonzomatamoros@aalto.fi

MÉTODO DE ELEMENTOS FINITOS MULTIESCALA PARA PROBLEMAS LINEALES

YESY KARINA SARMIENTO PERDOMO

Resumen. En este trabajo se estudia el método de elementos finitos multiescala combinado con técnicas de descomposición de dominio para la ecuación reacción-difusión en medios heterogéneos. El Método de Elementos Finitos Multiescala aprovecha las características de escala presentes en el problema para obtener soluciones precisas y de bajo costo computacional, en comparación al Método de Elementos Finitos clásico.

Abstract. This paper studies the multiscale finite element method combined with domain decomposition techniques for the reaction-diffusion equation in heterogeneous media. The Multiscale Finite Element Method takes advantage of the scale characteristics present in the problem to obtain precise solutions with low computational cost, compared to the classical Finite Element Method.

1. introducción

Los problemas multiescala en ciencia e ingeniería a menudo se describen mediante ecuaciones diferenciales parciales (PDEs) con coeficientes altamente oscilatorios. Las aplicaciones típicas incluyen flujos en medios porosos, problemas de transporte turbulento y la propagación de ondas sísmicas es una de las aplicaciones de mucho interés en la actualidad. Resolver estos problemas numéricamente es difícil porque una solución precisa generalmente requiere una resolución muy fina y, por lo tanto, una gran cantidad de memoria de computadora. La computación paralela reduce la dificultad hasta cierto punto, pero el tamaño de computación no se reduce a los enfoques tradicionales que resuelven directamente ecuaciones en mallas finas, es por ello que recientemente, se ha desarrollado un método de elementos finitos multiescala (MsFEM) [1] para capturar las soluciones a gran escala de problemas multiescala en una malla gruesa (con un tamaño de malla mayor que una determinada escala de corte del problema). La idea principal del método es construir la información local a pequeña escala del operador diferencial de orden principal en las funciones base de elementos finitos [3]. Es por ello que este trabajo tiene como objetivo principal estudiar el Método de Elementos Finitos Multiescala ya que puede ser útil para resolver problemas lineales que modelen un fenómeno físico que se necesite explicar, para beneficio de nuestra sociedad hondureña, ya sea para resolver un problema urgente o para beneficio del crecimiento de la ciencia en nuestro país.

La Universidad Autónoma de Honduras, dentro de los ejes de investigación que persigue, específicamente en el eje de Investigación 3: población y condiciones de vida,

Palabras clave. Método de Elementos Finitos Multiescala, Método de elementos finitos multiescala generalizado, medio heterogéneo.

enmarca como tema prioritario, el tema: cultura, ciencia y educación, situando este trabajo en el tema de Ciencia, ya que la matemática es naturalmente de carácter científico.

A continuación se presentan los antecedentes del método de elementos finitos multiescala donde se describe su trayectoria, así como su generalización que sucede posteriormente, luego, la justificación donde se plantean los argumentos del por qué es importante desarrollar el estudio del tema, se presenta el marco teórico, que es el que contiene todos los aspectos teóricos del método de elementos finitos multiescala y por último algunas conclusiones que surgieron al realizar el trabajo.

2. Antecedentes

El estudio del método de elementos finitos multiescala surge por primera vez en el año 1996, con el objetivo de resolver una clase particular de problemas elípticos, los cuales surgen de flujos en medios porosos y materiales compuestos [1]. En el año 1999, en [2] se propone un método de elementos finitos multiescala para resolver ecuaciones elípticas de segundo orden con coeficientes que oscilan rápidamente. Realizaron análisis de convergencia bajo el supuesto de que el coeficiente de oscilación es de dos escalas y periódico en escala rápida.

En el mismo trabajo se revela que el error de orden principal en este enfoque es ocasionado por el "muestreo resonante", el cual conduce a un gran error cuando el tamaño de la malla está cerca de la pequeña escala del problema contínuo. En [3] se plantea una técnica de sobremuestreo para hacerle frente a las dificultades presentadas en [2].

Años más tarde se propone una generalización de los métodos de elementos finitos multiescala (MsFEM) para problemas no lineales. Desarrollando estudios de la convergencia del método propuesto para ecuaciones elípticas no lineales y se propone una técnica de sobremuestreo [4]. En el año 2004, se desarrolla el método de elementos finitos multiescala para resolver numéricamente problemas escalares elípticos de valores en la frontera de segundo orden con coeficientes altamente oscilantes. Lo nuevo en este trabajo es que dicho método se basa en el acoplamiento de una malla global gruesa y una malla local fina, siendo esta última utilizada para calcular de forma independiente una base de elementos finitos adaptada para la malla gruesa [5].

Para el año 2010, se desarrolla un método de elementos finitos mixtos multiescala (MsMFE) para el modelado detallado de yacimientos vuggy, también conocidos como yacimientos cavernosos o yacimientos kársticos, son un tipo especial de yacimiento geológico caracterizado por la presencia de cavidades o espacios vacíos en la roca que los contiene. Este fue un primer paso hacia un marco multifísico multiescala uniforme.
En este estudio, las soluciones del método de elementos finitos multiescala se comparan con soluciones de Stokes-Brinkman a escala fina para casos de prueba que incluyen fracturas de corto y largo alcance [6].

Luego, en [7] para el año 2013, estudian MsFEM usando funciones base espectrales multiescala que están diseñadas para problemas de alto contraste. Las funciones de base multiescala se construyen usando vectores propios de un problema espectral local. La idea de este enfoque es mejorar la precisión y eficiencia de la simulación al capturar de manera eficiente las variaciones locales en las propiedades del material.

En [8] se propone un método de elementos finitos multiescala generalizado GMsFEM para la propagación de ondas elásticas en medios anisotrópicos heterogéneos, donde se construyen funciones base a partir de múltiples problemas locales tanto para los límites como para el interior de un soporte de nodo grueso.

En [9] se propone un marco de solución multiescala para el problema de equilibrio geomecánico de medios porosos heterogéneos basado en el método de elementos finitos. Después de imponer una cuadrícula de escala gruesa en el problema de escala fina dado, las funciones de base de escala gruesa se obtienen resolviendo problemas de equilibrio local dentro de elementos gruesos. Estas funciones básicas forman los operadores de restricción y prolongación utilizados para obtener el sistema de escala gruesa para el vector de desplazamiento.

Uno de los trabajos más recientes es [10], donde proponen un nuevo enfoque multiescala con una escala gruesa sin malla. Una escala gruesa se construye sobre la base de una cuadrícula computacional ya existente en una escala fina, dependiendo de los parámetros heterogéneos del problema. Este enfoque se basa en GMsFEM, donde los parámetros heterogéneos del problema se tienen en cuenta en una escala gruesa utilizando funciones de base multiescala. Estas funciones de base multiescala se construyen en una etapa fuera de línea utilizando problemas espectrales locales. Para representar las fracturas en una cuadrícula fina, se utiliza el Modelo de Fracturas Discretas.

De manera general los estudios han explorado el uso de métodos de elementos finitos multiescala no lineales, se ha revisado su aplicación en problemas con contrastes altos, también discutido sus fundamentos teóricos y aplicaciones en diversas áreas como el modelado de fractura en materiales heterogéneos.

3. Ecuación de Reacción-Difusión

En este trabajo se abordará la ecuación de reacción-difusión utilizando el método de elementos finitos multiescala, es por ello que se define a continuación dicha ecuación.

Sea

$$\frac{\partial u}{\partial t} - div(\kappa(x)\nabla u) = f(u),\ t > 0,\ x \in \mathbb{R}^n,$$

donde $u = u(x,t) \in \mathbb{R}$ representa la cantidad de una población, $t > 0$ el tiempo, $x \in \mathbb{R}^n$ la posición espacial y $\kappa(x)$ es la conductividad del medio. Esta ecuación es usada en la dinámica de cantidades físicas, biológicas o químicas.

Un caso particular de esta ecuación es el modelo de Fisher-KPP en 1937

$$\frac{\partial u}{\partial t} = \sigma u(1-u) + \kappa \frac{\partial^2 u}{\partial x^2},$$

donde $u = u(x,t)$ es la concentración de los miembros de una población distribuida uniformemente a lo largo de un hábitat lineal Ω (estructura del hábitat como ser un bosque).

Sea $q = q(x,t)$ la concentración de los miembros de la población cuyos descendientes tienen el gen mutante, el cual es asumido por $q = 1-u$. Luego σ denota la intensidad en favor del gen mutante independiente de u. Y supóngase que la razón por generación a la cual los miembros de la población con el gen mutante se difunde dentro de la población total está dada por $-\kappa \frac{\partial u}{\partial x}$, donde $\kappa > 0$ es una constante de difusión (independiente de x y u) [11].

3.1. **Condiciones de frontera y condición inicial.** Para garantizar la unicidad de soluciones en ecuaciones diferenciales se debe garantizar condiciones sobre la solución en el borde del dominio Ω, y/o condiciones iniciales que deben darse en un punto dado. Cuando se imponen condiciones sobre el borde del dominio Ω se tiene un problema de frontera; si las condiciones se dan en una subvariedad inicial se denomina un problema de Cauchy o de condición inicial.
Por ejemplo, considerando

$$\frac{\partial u}{\partial t} - div(\kappa(x)\nabla u) = f(u),\ t > 0,\ x \in \mathbb{R}^n$$

con condición inicial

$$u(x,0) = u_0(x),\ \text{para } x \in \Omega$$

y condiciones de frontera

$$b(x,t,u,\nabla u) = 0,\ \text{para } x \in \Omega,\ t > 0.$$

4. Método de elementos finitos multiescala

Los métodos de elementos finitos multiescala constan de dos ingredientes principales que son funciones de base multiescala y la formulación numérica global que combina estas funciones base multiescala. Las funciones base están diseñadas para capturar las características multiescala de la solución [12].

4.1. **Funciones base.** Sea τ^H una triangulación de la malla gruesa de Ω en elementos finitos (triángulos, cuadriláteros, etc.). Supongáse que la malla gruesa se puede resolver a través de una triangulación fina τ^h, como se ilustra en la Figura 1. Sean φ_i las funciones base del espacio de elementos finitos estándar $V^h = \mathbb{Q}^1(\tau^H)$ (que es el espacio de funciones bilineales por partes con respecto a la triangulación τ^H). La aproximación por elementos finitos estándar en la malla fina está dada como $u_h = \sum_i \alpha_i \varphi_i$, que genera el sistema lineal $Au_h = b$.

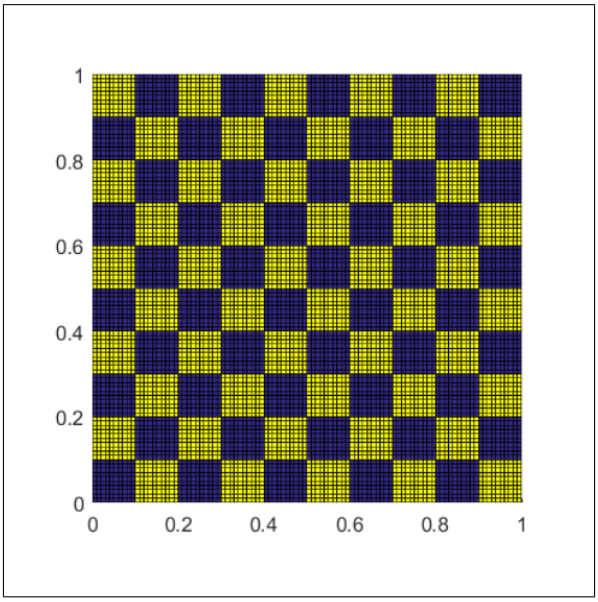


FIGURE 1. Cuadrícula de malla gruesa y fina.

Sea $R_0 = [\Psi_1, \Psi_2, ..., \Psi_N]$ una matriz (u operador) que transforma funciones definidas en la malla gruesa a funciones definidas en la malla fina. La cual está conformada por los vectores Ψ_1 (funciones base multiescala, las cuales son funciones de elementos finitos de la malla fina). Usando la aproximación:

$$u_h \approx R_0\delta, \text{ con } \delta \in \mathbb{R}^M, M \ll n \tag{1}$$

aquí M denota la dimensión del espacio de funciones que está asociado a la malla gruesa y n denota la dimensión del espacio de elementos finitos de la malla fina. Al remplazar (1) en el sistema lineal fino se obtiene que δ debe satisfacer el sistema sobredeterminado $AR_0\delta = \boldsymbol{b}$.
Para resolver el sistema se multiplica por el operador de escalamiento R_0^T (transforma funciones definidas en la malla fina a funciones definidas en la malla gruesa) a ambos lados de la igualdad para obtener $R_0^T AR_0\delta \approx R_0^T\boldsymbol{b}$. Al denotar $A_0 = R_0^T AR_0$ y $b_0 = R_0^T\boldsymbol{b}$ podemos obtener los coeficientes δ resolviendo el siguientes sistema lineal

$$A_0\delta = b_0.$$

De [7] se plantea una forma de construir las funciones base multiescala. Para introducir este método, sean $\{x_i\}_{i=1}^N$ los vértices de la malla gruesa τ^H y definimos la vecindad de los nodos x_i como

$$w_i = \bigcup\{E_j \in \tau^H | x_i \in \bar{E}_j\}$$

donde E_j son los elementos de la triangulación τ^H. Sea χ_i una familia de funciones que particiona la unidad subordinada a las vecindades w_i, una partición de la unidad es una técnica que se utiliza comúnmente en análisis y teoría de la medida para descomponer una función en una suma de funciones más simples y bien

comportadas. Se utiliza para extender la definición de una integral a espacios más generales que no son tan suaves o continuos. Ahora, se definen los vectores ψ_i de la matriz R_0 como:

$$\psi_{i,l} = \chi_i \varphi_l,$$

donde φ_l es una función con dominio w_i que representa la solución de malla fina.

Al desarrollar el método de elementos finitos multiescala generalizado, los vectores locales φ_l son aproximaciones del problema de valores y vectores propios generalizados:

$$-\text{div}(k\nabla\varphi) = \lambda k\varphi.$$

Ahora se complementa este problema de vectores propios con condiciones de frontera de Neumann, donde se usan solamente los vectores propios asociados a los menores valores propios λ_l. En [7] se muestra que el método de elementos finitos multiescala generalizado es adecuado para aproximar soluciones a la ecuacion de difusión en medios heterogéneos con alto contraste, en comparación con algunos otros métodos multiescala que no muestran buen desempeño para este tipo de problemas [12].

A continuación se desarrolla la formulación global del método de elementos finitos multiscala, una herramienta innovadora en la simulación de problemas complejos que abarcan múltiples escalas.

4.2. **Problema de reacción-difusión.** La representación de la solución a escala a través de funciones base multiescala permite reducir la dimensión del cálculo. Cuando la aproximación de la solución $p_h = \sum_i p_i \phi_i(x)$ puntos nodales de la cuadrícula. Se sustituyen en la ecuación de escala fina, el sistema resultante se proyecta en el espacio de dimensión gruesa para encontrar p_i. Esto se puede hacer multiplicando la ecuación de escala fina resultante con funciones de prueba de escala gruesa. Se pueden tomar otros enfoques para problemas generales no lineales [11].

En el caso de los métodos de elementos finitos de Galerkin, cuando las funciones base son conforme, es decir $\mathcal{P}_h \subset H_0'(\Omega)$, el MsFEM trata de encontrar $p_h \in \mathcal{P}_h$ tal que:

$$\sum_K \int_K k\nabla_{p_h}\nabla_{v_h} dx = \int_\Omega f v_h dx, \forall\ v_h \in \mathcal{P}_h. \tag{2}$$

Se pueden elegir las funciones de prueba de W_h (en lugar de $\mathcal{P}_h$) y llegar a la versión de Petrov-Galerkik de MsFEM, es decir
hallar $p_h \in \mathcal{P}_h$ tal que

$$\sum_K \int_K k\nabla_{p_h}\nabla_{v_h} dx = \int_\Omega f v_h dx, \forall\ v_h \in W_h. \tag{3}$$

Observe que en ambas formulaciones (2 y 3) el sistema de escala fina se multiplica por funciones de prueba de escala gruesa y por lo tanto el sistema resultante es de dimensiones gruesa. La ecuación (2) o (3) acopla las funciones de base multifocal. Esto da lugar a un sistema lineal de ecuaciones para encontrar los valores de la solución en los nodos del bloque de cuadrícula gruesa, por lo tanto, el sistema de ecuaciones lineales resultante determina la solución en la cuadricula gruesa [12].

En particular los elementos de la matriz de rigidez están dados por:

$$a_{ij} = \sum_k \int_k k\nabla\phi_i \nabla\phi_j^0 dx,$$

y se pueden aproximar usando:

$$\frac{1}{|k|}\int_k k\nabla\phi_i\nabla\phi_j^0 dx \approx \frac{1}{|k_{loc}|}\int_{k_{loc}} k\nabla\tilde{\phi_i}\nabla\phi_j^0 dx,$$

donde k_{loc} se refiere a la región computacional local y $\tilde{\phi_i}$ es la funcion base definida en k_{loc}.

La resolución de problemas multiescala que involucran coeficientes altamente contrastantes, representados por la función $\kappa(x)$, presenta desafíos significativos en la simulación numérica y el análisis de sistemas físicos y naturales. En estos casos, las variaciones abruptas en $\kappa(x)$ a lo largo del dominio pueden resultar en soluciones locales que cambian rápidamente y en cambios significativos en las propiedades globales del sistema, se abordará un ejemplo a continuación.

4.3. **Problema de Fisher en un medio con múltiples escalas y alto contraste.** Se planteará la formulación débil del problema de Fisher. Recalcamos que un coeficiente κ es conocido como coeficiente con múltiples escalas si tiene variaciones en todo el dominio a diferentes escalas. Por ejemplo, un coeficiente κ varía a escala ϵ si al tomar una región cualquiera de tamaño ϵ, se observan variaciones del coeficiente en esa región. El contraste se mide como el cociente entre el mayor valor y el menor valor del coeficiente de difusión (el cual es positivo).

Recordando que el problema a considerar es calcular $u : \Omega \to \mathbb{R}$ tal que:

$$\begin{cases} u_t = \sigma u(1-u) + \operatorname{div}(k(x)\nabla u), & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \Omega \times (0,T), \\ u(x,0) = u_0(x) \end{cases}$$

Se va a trabajar como un método iterativo, por lo que se procederá haciendo la sustitución de la forma: dado $u^{(0)}$ calcular, $u^{(1)}, u^{(2)}, ...$, al resolver,

$$\begin{cases} \frac{1}{\Delta t}u^{(n+1)} - \operatorname{div}(k(x)\nabla u^{(n+1)}) = \sigma u^{(n)}(1-u^{(n)}) + \frac{1}{\Delta t}u^{(n)} & \text{en } \Omega \\ u^{(1)} = 0, & x \in \partial\Omega \end{cases}$$

Ahora la formulación débil correspondiente es,

$$\begin{cases} \frac{1}{\Delta t}\int_\Omega u^{(n+1)}v + \int_\Omega k(x)\nabla u^{(n+1)}\nabla v = \int_\Omega [\sigma u^{(n)}(1-u^{(n)})]v + \dfrac{1}{\Delta t}\int_\Omega u^{(n)}v, \\ u(x,0) = u_0(x), \end{cases}$$

si se considera $V^h = \mathbb{P}^1(\tau^h)$ como el espacio de las funciones lineales por partes con respecto a la triangulación τ^h, la formulación de Galerkin genera un sistema

lineal

$$\mathbf{u}_h^{(n+1)} = \mathbf{b}^{(n+1)} \tag{4}$$

donde $\mathbf{b}^{(n+1)}$ depende de $\mathbf{u}_h^{(n)}$.

Una de las dificultades que se presenta al momento de resolver el sistema (4) es el tamaño del sistema lineal matricial, el cual es proporcional al tamaño de la malla (h^{-2}).
Para problemas multiescala con alto contraste en el coeficiente $\kappa(x)$, el número de condición de la matriz A depende del valor del contraste definido como $\eta = {}_{max}/_{min}$. El número de condición depende también de las variaciones locales del coeficiente de difusión κ en Ω. Además, se conoce que para el caso en que κ varía considerablemente se debe usar una malla muy fina, tan fina de modo que resuelva todas las variaciones del coeficiente κ. Esto conduce a que, en el caso en que κ tenga alto contraste y variaciones en varias escalas, el sistema lineal (4) es de dimensión muy grande y mal condicionado. Por lo que entonces el método de elementos finitos estándar aproxima una solución que debe ser calculada al resolver un sistema lineal con una matriz mal condicionada y de dimensión alta, con número de condición que depende de h^{-2} (donde h es el parámetro de tamaño de malla) y η es el contraste del coeficiente.

Dicho lo anterior, se debe buscar una alternativa que permita solucionar de manera eficiente estos sistemas lineales dispersos, de dimensión alta y mal condicionados, una alternativa para esta dificultad es elementos finitos multiescala, que consiste en proyectar el sistema lineal (4) a un sistema menor, generalmente asociado a una malla gruesa τ^H con tamaño que puede ser manejado de manera eficiente. La malla τ^H no necesita resolver todas las variaciones del coeficiente. Esto usualmente involucra la construcción de un operador de reducción de escala R_0, que transforma funciones definidas en la malla gruesa a funciones definidas en la malla fina y un operador de escalamiento R_0^T que transforma funciones definidas en la malla fina a funciones definidas en la malla gruesa. Al usar estos operadores, el sistema lineal (4) se convierte en un sistema lineal $A_0\delta = b_0$, de modo que $u_h = R_0$ se puedan calcular sin complicaciones [11].

Otra alternativa para el problema antes descrito es la construcción de un precondicionador usando técnicas de descomposición de dominios para solucionar el sistema (4). A continuación se brindan más detalles de este método.

5. Método de descomposición de dominio

Los métodos de descomposición de dominio se refieren a una colección de técnicas que consisten en dividir el dominio del problema en subdominios de tal manera que combinando soluciones de problemas en subdominios se puedan construir aproximaciones de las soluciones en el dominio original [10]. La idea original fue introducida por Hermann Schwarz, quien estaba interesado en la existencia y unicidad de la solución del problema de Poisson.

5.1. **Método alternado de Schwarz.** Para ilustrar la idea de este método, considere el dominio Ω formado por la unión de un círculo Ω_1 y un rectángulo Ω_2, de

esta forma, el problema de Poisson con condición de frontera de Dirichlet es:

$$\begin{cases} \Delta u = -f \text{ en } \Omega \\ u = g, \text{ en } \partial\Omega \end{cases} \tag{5}$$

La idea es dividir el problema de Poisson en

$$\begin{cases} \Delta_1 u = -f \text{ en } \Omega \\ u_1 = g_1, \text{ en } \partial\Omega_1 \end{cases} \tag{6}$$

y

$$\begin{cases} \Delta_2 u = -f \text{ en } \Omega \\ u_2 = g_2, \text{ en } \partial\Omega_2 \end{cases} \tag{7}$$

De tal forma que siguiendo estos cuatro pasos, Schwarz verificó convergencia en el dominio Ω:

(1) Resolver (6) al completar g_1 arbitrariamente en $\partial\Omega_1$.
(2) Resolver (7) al usar la solución del paso (1) para completar los datos de frontera.
(3) Resolver (6) al usar la solución de (5) para completar el dato de frontera.
(4) Iterar hasta convergencia.

5.2. **Método en paralelo de Schwarz.** Otro método que implementó Schwarz para resolver problemas con uniones de geometrías simples es el denominado método en paralelo de Schwarz. Este método consiste en resolver iteradamente los siguientes dos pasos:

(1) Resolver los problemas de Poisson (6) y (7) al mismo tiempo en paralelo de tal forma que las soluciones u_1 y u_2 respectivamente de los problemas se tiene que

$$v = B_1 u_1 + B_2 u_2.$$

Aquí, B_1 y B_2 son extensiones por cero fuera del dominio Ω_1 y Ω_2, respectivamente, las "extensiones por cero" son una técnica que se utiliza para extender una función definida en un subconjunto de un dominio más grande a una función en todo el dominio, manteniendo ciertas propiedades.
(2) Implementar nuevamente (1) y usar v para completar datos de frontera. Al aplicar el método de elementos finitos a los problemas (6) y (7), las soluciones numéricas están dadas por los sistemas $A_1\alpha_1 = b_1$ y $A_2\alpha_2 = b_2$ respectivamente.

Para las soluciones α_1 y α_2 se define el primer paso de el método en paralelo de Schwarz, esto es, $v = B_1\alpha_1 + B_2\alpha_2$. Observe que al definir los sistemas lineales de los problemas (6) y (7), de la forma $\alpha_1 = A_1^{-1}b_1$ y $\alpha_2 = A_2^{-1}b_2$ respectivamente, y multiplicar B_i^T (operador de restricción) se tiene

$$v = (B_1 A_1^{-1} + B_2 A_2^{-1} B_2^T)b.$$

Se puede considerar entonces $M^{-1} = B_1 A_1^{-1} B_1^T + B_2 A_2^{-1} B_2^T$ como un precondicionador de la matriz A.

La construcción del precondicionador permite retomar la idea en resolver el sistema obtenido a través del siguiente sistema con precondicionador

$$M^{-1}A\mathbf{u}_h^{n+1} = M^{-1}\mathbf{b}^{n+1} \tag{8}$$

donde la complejidad computacional depende ahora de cond$(M^{-1}A)$ (condición de la matriz $M^{-1}A$) y se espera que cond$(M^{-1}A)$ sea mejor que la condición de la matriz A. Para resolver el sistema (8) con precondicionador, aplicado a través del método en paralelo de Schwarz, se hace necesario un método iterativo que permita resolver los sistemas resultantes [5].

Por otro lado, al considerar el método en paralelo de Schwarz para varios subdominios, el precondicionador se define como $M_1^{-1}b = \sum_{i=1}^{N} B_i A_i^{-1} B_i^T b$. Esta condición es conocida como método aditivo de un nivel de Schwarz. Es de mencionar que entre más subdominios se presenta más lenta es la convergencia. Al considerar la malla gruesa del dominio para obtener un mejor precondicionador, se tiene

$$M_2^{-1}b = M_1^{-1}b + R_0 A_0^{-1} R_0^T b$$

donde el segundo término está asociado a la corrección dada por la malla gruesa. Esta última parte se puede definir por el método de elementos finitos multiescala, a esto se le llama método aditivo de dos niveles de Schwarz. Teniendo en cuenta este último método, se determina que para problemas elípticos

$$\text{cond}(M^{-1}A) \leq C\Lambda\left(1 + \frac{H^2}{\hat{\delta}^2}\right)$$

la cual C es una constante, $\hat{\delta}$ es un subreposición, H es el diámetro de la vecindad w_i y

$$\Lambda = \max_{1\leq i\leq N_s} \frac{1}{\lambda_{L_{i+1}}}$$

donde en la vecindad w_i se usan los vectores propios $\{\varphi_l\}$ asociados a los menores valores propios $\{\lambda_l\}$ para $1 \leq l \leq L_i$. Observe que $C\Lambda$ no depende del contraste η si se toma suficientes valores propios [12].

6. Conclusiones

En este trabajo se exponen métodos iterativos que combinan ideas de descomposición de dominio con ideas de elementos finitos multiescala para la solución de la ecuación de reacción-difusión.

Estos métodos han sido expuestos en [11] y [12] para mostrar como estas técnicas pueden ser usadas para aliviar el tiempo computacional de una ecuación de reacción-difusión en medios heterogéneos con múltiples escalas y alto contraste.

En este trabajo también se presentaron de forma detallada los conceptos relacionados a la ecuación de reacción-difusión, los conceptos sobre elementos finitos multiescala y descomposición de dominio, con el objetivo de darle solución a los problemas que se presentan a la hora de resolver esta ecuación.

Un detalle importante presentado por [11] es la complicación que las variaciones

del coeficiente $\kappa(x)$ y el contraste presente en el mismo afectan negativamente el núumero de condición de este sistema lineal obtenido. Por lo que se requieren técnicas iterativas como las expuestas en este trabajo para resolver estos sistemas de ecuaciones.

Este trabajo se puede considerar como punto de partida para una futura investigación donde se hagan experimentos numéricos para comprobar la eficiencia de los métodos, para ello se sugiere realizar implementaciones de estos métodos en un lenguaje de programación.

References

1. Hou, T. Y., Wu, X.-H. (1996). *A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media.* Applied Mathematics, Caltech, Pasadena, California 91125.
2. Hou, T. Y., Wu, X.-H., Cai, Z. (1999). *Convergence of a Multiscale Finite Element Method for Elliptic Problems with Rapidly Oscillating Coefficients.*
3. Efendiev, Y. R., Hou, T. Y., Wu, X.-H. (2000). *Convergence of a Nonconforming Multiscale Finite Element Method.*
4. Efendiev, Y., Hou, T., Ginting, V. (2004). *Multiscale Finite Element Methods for Nonlinear Problems and Their Applications.*
5. Allaire, G., Brizzi, R. (2004). *A Multiscale Finite Element Method for Numerical Homogenization.* Centre de mathématiques appliquées, France.
6. Gulbransen, A. F., Hauge, V. L., Lie, K.-A. (2008). *A Multiscale Mixed Finite-Element Method for Vuggy and Naturally-Fractured Reservoirs.* Department of Applied Mathematics, SINTEF ICT.
7. Efendiev, Y., Galvis, J., Wu, X.-H. (2013). *Multiscale Finite Element Methods for High-Contrast Problems Using Local Spectral Basis Functions.*
8. Gao, K., Fu, S., Gibson Jr., R. L., Chung, E. T., Efendiev, Y. (2015). *Generalized Multiscale Finite-Element Method (GMsFEM) for Elastic Wave Propagation in Heterogeneous, Anisotropic Media.*
9. Castelletto, N., Hajibeygi, H., Tchelepi, H. A. (2016). *Multiscale Finite-Element Method for Linear Elastic Geomechanics.*
10. Fu, S., Altmann, R., Chung, E. T., Maier, R., Peterseim, D., Pun, S.-M. (2016). *Meshfree Generalized Multiscale Finite Element Method.* Department of Mathematics, University of Augsburg, Germany.
11. Hernández, J. D. (2018). *Método de Elementos Finitos Para la Ecuación de Reacción-Difusión en Medios Heterogéneos.* Boletín de Matemáticas 25(2) 123–138.
12. Efendiev, Y., Hou, T. (2008). *Multiscale Finite Element Methods. Theory and Applications.*

Dirección: Escuela de Matemática, Universidad Nacional Autónoma de Honduras,Tegucigalpa, Honduras.

Correo: `yesy.sarmiento@unah.hn`

VISIÓN ALGEBRAICA: BASES DE GRÖBNER EN VISIÓN COMPUTACIONAL

LEONEL OBANDO

Abstract. En este artículo se estudiarán los fundamentos teóricos para enfrentar problemas de visión computacional utilizando enfoques algebraicos. El estudio se centra en las bases de Gröbner para ideales generados por los polinomios involucrados en un sistema de ecuaciones polinomiales. Se propone el algoritmo de Buchberger clásico para el cálculo de estas bases y se propone cómo obtener una versión reducida de estas bases.

This article will study the theoretical foundations for addressing computer vision problems using algebraic approaches. The study focuses on Gröbner bases for ideals generated by the polynomials involved in a system of polynomial equations. The classic Buchberger algorithm is proposed for computing these bases, along with a method for obtaining a reduced version of these bases.

.

1. Introducción

Las computadoras, desde su creación, se han utilizado para resolver, con mayor eficiencia, problemas o tareas realizadas anteriormente por el ser humano. Los constantes y acelerados avances tecnológicos han logrado que las computadoras puedan realizar trabajos específicos más allá de ejecutar de cálculos complejos con precisión. Por ejemplo, la robótica ha intentado emular el movimiento y la forma humana, la inteligencia artificial busca imitar los procesos cognitivos del ser humano y la *visión computacional* tiene como objetivo reproducir el proceso de visión humana. Más específicamente, en visión computacional se estudia cómo las computadoras pueden adquirir, procesar, analizar y entender la información contenida imágenes en 2D del mundo real de manera muy similar a como lo hacemos los humanos con el mínimo esfuerzo cada día. Algunas aplicaciones en radiación, procesamiento de señales, reconocimiento de patrones pueden encontrarse en [9].

Muchos de los problemas de visión computacional involucran la solución de un sistema de ecuaciones polinomiales. Estos sistemas pueden resolverse utilizando métodos iterativos como el conocido método de Newton. Sin embargo, los métodos iterativos están ligados a un determinado orden de convergencia y a adecuada elección de la aproximación inicial. Una alternativa a estos métodos consiste en eliminar variables del sistema utilizando métodos algebraicos provenientes de geometría algebraica. A esta fusión entre la visión computacional y la geometría

Fecha: 21 de agosto de 2023.
Palabras y frases clave. Visión Computacional, Visión algebrica, Bases de Gröbner, Algoritmo de Buchberger.

algebraica se le conoce como *visión algebriaca*.

La técnica específica de geometría algebrica para resolver sistemas de ecuaciones polinomiales que forma parte principal de este artículo consiste en las bases de Gröbner tienen aplicaciones en una variedad de campos, incluida la geometría algebraica y la visión computacional [3]. En esencia, una base de Gröbner es un conjunto especial de polinomios que puede representar y describir completamente un ideal en un anillo de polinomios. La principal característica de una base de Gröbner es que tiene la propiedad única de transformar cualquier polinomio perteneciente al ideal original en una forma simplificada y específica. El proceso de cálculo de bases de Gröbner involucra el algoritmo de Buchberger, que toma un conjunto de polinomios generadores de un ideal y genera iterativamente nuevos polinomios que forman la base de Gröbner. Estos polinomios ayudan a simplificar los cálculos y a analizar propiedades algebraicas.

En este artículo se estudia en profundidad las bases de Gröbner, sus propiedades principales, el algoritmo de Buchberger para calcular bases de Gröbner con el objetivo de comprender una de las técnicas algebricas más utilizadas en visión computacional.

2. Justificación

El potencial de las técnicas de visión computacional se manifiesta en una gran variedad de áreas. Se han utilizado tecnologías de reconocimiento óptico de caracteres (OCR) para la identificación de vehículos mediante sus placas en [13]. Similarmente, ramas de la medicina como cardiología, patología, dermatología y oftalmología se han visto beneficiadas de la combinación de inteligencia artificial con visión computacional [6]. Recientemente, la visión computacional ha servido como un pilar para el desarrollo de vehículos autónomos como el caso de vehículos aéreos no tripulados (UAV) [10]. Por lo tanto, el desarrollo de métodos que permitan resolver problemas de visión computacional puede significar avances tecnológicos interesantes en diferentes áreas de la ciencia.

Dentro de las líneas de investigación de la Universidad Nacional Autónoma de Honduras (UNAH), este artículo puede ser ubicado en el eje de investigación *Población y Condiciones de Vida* dentro del tema prioritario *Cultura, ciencia y educación*. Esto debido a que el objetivo del presente estudio es la divulgación científica. Así mismo, los problemas de visión computacional que se buscan resolver con el contenido de este artículo pertenecen a la línea de investigación de *Modelación Matemática* de la Orientación en Ingeniería Matemática de la Maestría en Matemática de la UNAH.

3. Antecedentes

El desarrollo de las bases de Gröbner ha marcado un hito fundamental en el campo del álgebra computacional, transformando la manera en que abordamos problemas algebraicos y geométricos. Desde su concepción en 1965 por el matemático Bruno Buchberger [2], las bases de Gröbner han evolucionado desde ser una idea innovadora hasta convertirse en una herramienta esencial con aplicaciones que se

extienden a una variedad de disciplinas, incluida la visión computacional.

En sus orígenes, Bruno Buchberger propuso las bases de Gröbner como una manera sistemática de abordar la teoría de ideales y resolver sistemas de ecuaciones polinomiales. Su algoritmo, llamado Algoritmo de Buchberger proporcionó un método efectivo para calcular estas bases, que se convirtieron en un marco unificador para entender la estructura algebraica subyacente de los ideales polinomiales. Durante la década de 1970, estas bases comenzaron a revelar su potencial en el ámbito de la geometría algebraica y el álgebra conmutativa [4].

A medida que avanzaba la década de 1980, los investigadores como H. Möller y Teo Mora [12] contribuyeron a la refinación y mejora de los algoritmos de cálculo de bases de Gröbner. Estos algoritmos permitieron manejar sistemas de ecuaciones más grandes y complejos, ampliando las aplicaciones en la geometría algebraica, la resolución de sistemas de ecuaciones y la teoría de códigos [8]. En la década de 1990, la intersección de las bases de Gröbner con la visión computacional comenzó a vislumbrarse, ya que problemas de correspondencia de puntos y reconstrucción 3D requerían un enfoque algebraico [17].

La década de 2000 marcó un período de expansión para las bases de Gröbner, ya que encontraron aplicaciones en áreas como la criptografía [1] y la resolución de sistemas polinomiales en campos finitos [7]. Investigadores como Bernd Sturmfels [16] y Michael Stillman [14, 15] jugaron un papel clave en la implementación y optimización de algoritmos relacionados con las bases de Gröbner. A medida que avanzamos en el siglo XXI, las bases de Gröbner se mantienen relevantes en la resolución de problemas algebraicos y geométricos, incluso en aplicaciones interdisciplinarias.

Hoy en día, las bases de Gröbner siguen influyendo en diversos campos, y su intersección con la visión computacional ha permitido abordar problemas complejos en la percepción visual. Investigadores como Zuzana Kukelova [11] han demostrado cómo las bases de Gröbner pueden aplicarse en la calibración de cámaras y la reconstrucción tridimensional, lo que muestra cómo esta herramienta algebraica ha evolucionado para abordar desafíos en la resolución de sistemas polinomiales en contextos prácticos.

4. Preliminares

Para estudiar las bases de Gröbner, es indispensable tener una comprensión sobre la teoría algebraica de polinomios. Debemos comenzar definiendo qué es un monomio.

Definición 4.1 (**Monomio**)**.** Un monomio en las variables $x_1, x_2, \ldots, x_n$ es un producto de la forma

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \tag{4.1}$$

donde todos los exponentes $\alpha_1, \alpha_2, \ldots, \alpha_n$ son enteros no negativos. El **grado total** de este monomio es la suma $\alpha_1 + \alpha_2 + \cdots + \alpha_n$.

Podemos simplificar la notación escribiendo $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ y así

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

Definición 4.2 (**Polinomio**)**.** Un polinomio f en las variables $x_1, x_2, \ldots, x_n$ con coeficientes en un campo k es una combinación lineal finita (con coeficientes en k) de monomios. Escribiremos entonces:

$$f = \sum_\alpha a_\alpha x^\alpha, \quad a_\alpha \in k \tag{4.2}$$

donde la suma es sobre todas las n-tuplas $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. El conjunto de todos los polinomios en $x_1, \ldots, x_n$ se denota como $k[x_1, \ldots, x_n]$ y lo llamaremos *anillo polinomial*.

Definición 4.3. Sea $f = \sum_\alpha a_\alpha x^\alpha$ un polinomio en $k[x_1, \ldots, x_n]$

i) Llamamos a a_α el coeficiente del monomio x_α.
ii) Si $a_\alpha \neq 0$, llamamos a $a_\alpha x^\alpha$ un término de f.
iii) El grado total de $f \neq 0$, denotado por $\deg(f)$, es el máximo $|\alpha|$ tal que el coeficiente a_α es no nulo.

Definición 4.4 (**Variedad Afín**)**.** Sea k un campo y sean $f_1, f_2, \ldots, f_s$ polinomios en $k[x_1, \ldots, x_n]$. Definamos

$$\mathbf{V}(f_1, \ldots, f_s) = \{(a_1, \ldots, a_n) \in k^n |\ f_i(a_1, \ldots, a_n) = 0, \text{ para cada } 1 \leq i \leq s\} \tag{4.3}$$

como la variedad afín definida por $f_1, \ldots, f_s$.

En otras palabras, una variedad afín no es más que el conjunto de soluciones del sistema polinomial

$$\begin{cases} f_1(x_1, \ldots, x_n) &= 0 \\ & \vdots \\ f_s(x_1, \ldots, x_n) &= 0 \end{cases} \tag{4.4}$$

Definición 4.5 (**Ideal Generado**)**.** Sean $f_1, f_2, \ldots, f_s$ polinomios en $k[x_1, \ldots, x_n]$. Llamaremos ideal generado por $f_1, f_2, \ldots, f_s$ a

$$\langle f_1, f_2, \ldots, f_s \rangle = \left\{ \sum_{i=1}^{s} h_i \mid h_1, \ldots, h_s \in k[x_1, \ldots, x_n] \right\} \tag{4.5}$$

5. Orden Monomial

El propósito de esta sección es sentar las bases para extender el algoritmo de la división de polinomios en $k[x]$ en el cual, dados los polinomios $f, f \in k[x]$ con $g \neq 0$ existen $q, r \in k[x]$ tales que

$$f(x) = q(x)g(x) + r(x)$$

donde $\deg(r) < \deg(g)$. Esta última condición implícitamente está imponiendo un orden en los polinomios de una variable. En esta sección lo haremos de forma explícita para monomios en general y, por extensión, a polinomios en general.

Definición 5.1 (**Orden Monomial**)**.** Un orden monomial $>$ para el conjunto de los monomios $\mathbf{x}^\alpha$ es un orden total que satisface

1) Si $\mathbf{x}^\alpha > \mathbf{x}^\beta$ y $\mathbf{x}^\gamma$ es cualquier monomio, entonces $\mathbf{x}^\alpha\mathbf{x}^\gamma > \mathbf{x}^\beta\mathbf{x}^\gamma$.
2) El conjunto de monomios con el orden $>$ es un conjunto bien ordenado.

Los órdenes monomiales más frecuentemente utilizados son los sigientes:

Definición 5.2 (**Orden Lexicográfico**)**.** Diremos que $\mathbf{x}^\alpha >_{lex} \mathbf{x}^\beta$ cuando en $\alpha - \beta \in \mathbb{Z}^n$ la primera entrada no nula (leyendo de izquierda a derecha) sea positiva.

Definición 5.3 (**Orden Lexicográfico Graduado**)**.** Diremos que $\mathbf{x}^\alpha >_{glex} \mathbf{x}^\beta$ cuando

$$|\alpha| > |\beta| \text{ o bien } |\alpha| = |\beta| \text{ pero } x^\alpha >_{lex} x^\beta$$

Es decir, el orden está determinado por el grado del monomio en primer lugar y en caso de empate se resuelve con el orden lexicográfico usual.

Definición 5.4 (**Orden Lexicográfico Inverso Graduado**)**.** Diremos que $\mathbf{x}^\alpha >_{grevlex} \mathbf{x}^\beta$ cuando $|\alpha| > |\beta|$ o bien $|\alpha| = |\beta|$ y la primera entrada no nula (leyendo de derecha a izquierda) de $\alpha - \beta \in \mathbb{Z}^n$ sea negativa.

Con estas opciones de órdenes monomiales, la sigiuente tarea consiste en poder ordenar los términos de un polinomio dado.

Ejemplo 5.5. Dado el polinomio $f = 4xy^2z + 4z^2 - 5x^3 + 7x^2z^2$ podemos ordenarlo de diferentes formas según el orden elegido:

Orden Lexicográfico: $f = -5x^3 + 7x^2z^2 + 4xy^2z + 4z^2$

Orden Lexicográfico Graduado: $f = 7x^2z^2 + 4xy^2z - 5x^3 + 4z^2$

Orden Lexicográfico Inverso Graduado: $4xy^2z + 7x^2z^2 - 5x^3 + 4z^2$

Con el orden definido, podemos definir la siguiente terminología:

Definición 5.6. Sea $f = \sum_\alpha a_\alpha x^\alpha$ un polinomio no nulo en $k[x_1, \ldots, x_n]$ y sea $>$ un orden monomial.

i) El **Término Principal** de f es $LT(f) = a_{\hat{\alpha}} x^{\hat{\alpha}}$ donde $x^{\hat{\alpha}}$ es monomio más grande que aparece en f con respecto a $>$.

ii) El **Coeficiente principal** de f, denotado $LC(f)$, es el coeficiente del término principal.
iii) El **Monomio Principal** de f es $LM(f) = LT(f)/LC(f)$

6. Bases de Gröbner

Con lo anterior expuesto podemos enunciar el siguiente teorema cuya demostración se encuentra en [5]

Teorema 6.1 (Algoritmo de la División)**.** *Sea $>$ algún orden monomial en $k[x_1, \ldots, x_n]$ y sea $F = (f_1, \ldots, f_m)$ un conjunto ordenado de polinomios en $k[x_1, \ldots, x_n]$. Luego, todo polinomio $f \in k[x_1, \ldots, x_n]$ puede escribirse como*

$$f = q_1 f_1 + q_2 f_2 + \cdots + q_m f_m + r \tag{6.1}$$

donde $q_i, r \in k[x_1, \ldots, x_n]$, $i = 1, 2, \ldots, m$ son polinomios tales que $LT(f) \geq LT(f_i q_i)$ con respecto a $>$ y $r = 0$ o bien, es una combinación lineal de monomios que no son divisibles por ninguno de los $LT(f_i)$, $i = 1, 2, \ldots, m$. Llamaremos a r el residuo de la división por F y lo denotaremos como $r = \overline{f}^F$.

La necesidad de que el conjunto de polinomios F sea ordenado radica en que para llevar a cabo este algoritmo se debe empezar haciendo divisiones sucesivas entre f_1 hasta obtener q_1 de modo que $LT(f - f_1 q_1)$ no sea divisible entre $LT(f_1)$. Se repite el procedimiento con los demás polinomios de F en ese orden especfíco. Resulta que el orden en el que estén las funciones puede alterar el residuo obtenido.

Ejemplo 6.2. Sean $f = x^2y + xy^2 + y^2$, $f_1 = y^2 - 1$ y $f_2 = xy - 1$. Con el orden lexicográfico con $x > y$

i) Si $F = (f_1, f_2)$, obtenemos que $\overline{f}^F = 2x + 1$ ya que
$$x^2y + xy^2 + y^2 = (x + 1)(y^2 - 1) + x(xy - 1) + 2x + 1$$
ii) Si $F = (f_2, f_1)$, obtenemos que $\overline{f}^F = x + y + 1$ ya que
$$(x + y)(xy - 1) + 1 \cdot (y^2 - 1) + x + y + 1$$

La existencia de las bases de Gröbner para cualquier ideal I de $k[x_1, \ldots, x_n]$ comienza en que dicho ideal debe estar generado por un conjunto finito de polinomios. Esto lo garantiza el sigiuente teorema demostrado en [5]:

Teorema 6.3 (Teorema de la Base de Hilbert)**.** *Cada ideal $I \subset k[x_1, \ldots, x_n]$ tiene un conjunto generador finito. Es decir, $I = \langle g_1, \ldots, g_t \rangle$ para algunos $g_1, \ldots, g_t \in I$.*

Sabiendo esto, a todo ideal polinomial puede calculársele una base de Gröbner la cual se define como:

Definición 6.4 (**Bases de Gröbner y Algoritmo de Buchberger**)**.** Sea I un ideal y $>$ un orden monomial en $k[x_1, x_2, \ldots, x_n]$. Una base de Gröbner para I con respecto a $>$ es un conjunto de polinomios $G = \{g_1, g_2, \ldots, g_l\}$, $G \subset I$ con la propiedad de que para cada polinomio no nulo $f \in I$, LT(f) es divisible por LT(g_i)

para algún i.

Estas bases de Gröbner son importantes debido a las sigientes dos propiedades demostradas en [5]:

1) Sea $I \subset k[x_1 \ldots, x_n]$ un ideal y sea $G = \{g_1, \ldots, g_t\}$ una base de Gröbner para I. Entonces dado $f \in k[x_1, \ldots, x_n]$, existe un único $r \in k[x_1, \ldots, k_n]$ tal que:
   a) Ningún término de r es divisible por alguno de $LT(g_1), \ldots, LT(g_t)$.
   b) Existe un polinomio $g \in I$ tal que $f = g + r$.

ii) Un polinomio $f \in I$ si y solo si $\overline{f}^G = 0$.

La primera de las propiedades destaca que, en el caso particular que $F = G$ en el Teorema 6.1, no importa el orden en el que se coloquen los elementos de G siempre se obtendrá el mismo residuo.

Definiremos ahora la última herramienta que nos permitirá construir el algoritmo de Buchberger.

Definición 6.5. Sean $f, g \in k[x_1, \ldots, x_n]$ polinomios no nulos.

i) Sean $LM(f) = x^\alpha$ y $LM(g) = x^\beta$ y sea $\gamma = (\gamma_1, \ldots, \gamma_n)$ donde $\gamma_i = \max(\alpha_i, \beta_i)$ para cada i. Llamaremos a x^γ el **Mínimo Común Múltiplo** de $LM(f)$ y $LM(g)$.
ii) El **S-polinomio** de f y g es la combinación lineal

$$S(f,g) = \frac{x^\gamma}{LT(f)} f - \frac{x^\gamma}{LT(g)} g \tag{6.2}$$

donde x^γ es el mínimo común múltiplo de f y g.

El objetivo del S-polinomio es obtener un polinomio a partir de f y g que elimine sus términos principales y se obtenga así un polinomio más pequeño en el orden especificado.

El criterio más importante para saber si un conjunto de polinomios dado es una base de Gröbner se encuentra en el siguiente teorema demostrado en [5]:

Teorema 6.6 (Criterio de Buchberger)**.** *Sea I un ideal de polinomios. Un conjunto finito de polinomios $G = \{g_1, g_2, \ldots, g_l\}$, $G \subset I$ es una base de Gröbner de I si y solo si $\overline{S(f,g)}^G = 0$ para todos los pares $i, j \in \{1, 2, \ldots, l\}$, con $i \neq j$.*

El siguiente algoritmo permite calcular una base de Gröbner a partir de un conjunto inicial de polinomios $F = \{f_1, \ldots, f_m\}$.

Algorithm 1: Algoritmo de Buchberger

Data: $F = \{f_1, \ldots, f_m\}$
Result: Una base de Gröbner $G = \{g_1, \ldots, g_t\}$ para $I = \langle f_1, \ldots, f_m \rangle$ con $F \subset G$

```
G = F;
do
    G' = G;
    for cada par p, q tal que g_p, g_q ∈ G' con p ≠ q do
        S = S(g_p, g_q)^{G'};
        if S ≠ 0 then
            G = G ∪ {S};
        end
    end
while G = G';
```

En este algoritmo se inicia con un conjunto de polinomios $f_1, \ldots, f_m$, los cuales perfectamente podrían ser polinomios que definen un sistema polinomial como en la ecuación (4.4). Se empieza poniendo como candidato a la base de Gröbner al propio conjunto de polinomios $G = \{f_1, \ldots, f_m\}$. Posteriormente se calculan los S-polinomios entre cada par de elementos de la propuesta base. Si el S-polinomio no deja residuo cero al dividirse entre G, entonces se añade el S-polinomio a la base y se repite todo el proceso hasta que todos los S-polinomios dejan residuo cero al dividirse entre G. Se garantiza que el algoritmo termina en una cantidad finita de pasos por el Teorema de la Base de Hilbert. El ideal generado por $f_1, \ldots, f_m$ debe de tener una base de Gröbner finita.

Sin embargo, tenemos el problema de que probablemente la base de Gröbner resultante del algoritmo sea bastante grande ya que incluye al conjunto F. Los siguientes resultados, demostrados en [5], proponen un método de reducir el tamaño de estas bases de Gröbner y que la versión más reducida posible es única para cada ideal.

Teorema 6.7. *Sea G una base de Gröbner de $I \subset k[x_1, \ldots, x_n]$. Sea $p \in G$ un polinomio tal que $LT(p) \in \langle LT(G) \rangle$. Entonces $G \backslash \{p\}$ es también una base de Gröbner de I.*

Podemos aplicar iterativamente este teorema hasta obtener:

Definición 6.8 (**Base de Gröbner Reducida**)**.** Una base de Gröbner reducida para un ideal polinomial I es una base de Gröbner G para I tal que

i) $LC(p) = 1$ para todo $p \in G$.
ii) Para todo $p \in G$, ningún monomio de p está en $\langle LT(G \backslash \{p\}) \rangle$.

Teorema 6.9. *Sea $I \neq \{0\}$ un ideal polinomial. Entonces, dado un orden monomial, I tiene una base de Gröbner reducida. Más aun, esta base de Gröbner reducida es única.*

7. Conclusiones y Trabajos Futuros

A lo largo de este artículo se ha podido evidenciar el potencial que tienen las bases de Gröbner cuando de resolver sistemas de ecuaciones polinomiales se trata. El algoritmo de Buchberger para calcular bases de Gröbner y el Teorema 6.9 nos permite partir de un sistema de ecuaciones polinomiales y, en lugar de buscar directamente la solución, estudiamos el ideal generado por dichos polinomios para encontrar la base de Gröbner reducida del ideal. Nos quedará así un sistema polinomial mucho más sencillo que el original.

Sin embargo, el algoritmo de Buchberger realiza un número considerable de iteraciones. Se estudiará cuál es el orden del número de operaciones realizadas por este algoritmo. Además, se estudiarán versiones mejoradas del algoritmo y se buscará realizar implementaciones en Python, Matlab o Mathematica para realizar experimentos numéricos y evaluar el desempeño de las diferentes variantes del algoritmo de Buchberger.

Por otro lado, se hará un estudio más profundo de las técnicas de inteligencia artificial que utilizan este algoritmo en problemas de reconstrucción de imágenes 3D a partir de un determinado número de imágenes 2D. O bien, en problemas de calibración de cámaras como los planteados en [11].

References

1. Ackermann, Peter, and Martin Kreuzer. *Gröbner basis cryptosystems.* Applicable Algebra in Engineering, Communication and Computing 17 (2006): 173-194.
2. Buchberger, Bruno. *Bruno Buchberger's PhD thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal.* Journal of symbolic computation 41.3-4 (2006): 475-511.
3. Buchberger, Bruno, and Franz Winkler, eds. *Gröbner bases and applications.* Vol. 251. Cambridge University Press, 1998.
4. Buchberger, Bruno. *Some properties of Gröbner-bases for polynomial ideals.* ACM SIGSAM Bulletin 10.4 (1976): 19-24.
5. Cox, David, John Little, and Donal OShea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra.* Springer Science & Business Media, 2013.
6. Esteva, Andre, et al. *Deep learning-enabled medical computer vision.* NPJ digital medicine 4.1 (2021): 5.
7. Gao, Sicun. Counting zeros over finite fields with Gröbner bases. Diss. Master's thesis, Carnegie Mellon University, 2009.
8. Imai, Hideki. *Multivariate polynomials in coding theory.* International Conference on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984.
9. Jahne, B. y Haussecker, H. *Computer vision and applications: a guide for students and practitioners.* Elsevier, 2000.
10. Kakaletsis, Efstratios, et al. *Computer vision for autonomous UAV flight safety: An overview and a vision-based safe landing pipeline example.* Acm Computing Surveys (Csur) 54.9 (2021): 1-37.
11. Kukelova, Zuzana. *Algebraic methods in computer vision.* Diss. Czech Technical University, 2013.
12. H.M. Möller, F. Mora, *Upper and Lower Bounds for the Degree of Groebner Bases*, in EUROSAM 1984, ed. by J. Fitch. Lecture Notes in Computer Science, vol. 174 (Springer, New York-Berlin-Heidelberg, 1984), pp. 172–183
13. Lotufo, R.A., Morgan, A.D. y Johnson, A.S, *Automatic number-plate recognition.* IEE Colloquium on Image Analysis for Transport Applications, London, UK, 1990, pp. 6/1-6/6.

14. D. Bayer, M. Stillman, *On the Complexity of Computing Syzygies, in Computational Aspects of Commutative Algebra*, ed. by L. Robbiano (Academic Press, New York, 1988), pp. 1–13
15. D. Grayson, M. Stillman, *Macaulay2, a Software System for Research* (2013), version 1.6, available at http://www.math.uiuc.edu/Macaulay2/
16. L. Pachter, B. Sturmfels (eds.), *Algebraic Statistics for Computational Biology* (Cambridge University Press, Cambridge, 2005)
17. Werman, R., and Amnon Shashua. *The study of 3D-from-2D using elimination*. Proceedings of IEEE International Conference on Computer Vision. IEEE, 1995.

Universidad Nacional Autónoma de Honduras
Dirección de correo electrónico: leonel.obando@unah.edu.hn

INTRODUCCIÓN A LA TEORÍA DE CURVAS ELÍPTICAS Y SU USO EN CRIPTOGRAFÍA

CARLOS URRUTIA

Abstract. Since the introduction of the public key concept in cryptography, the potential to use the discrete logarithm problem has been developed, which is defined as the problem of calculating logarithms with respect to the generator on a multiplicative group of modulo prime integers. The ideas proposed to address this problem can be extended to arbitrary groups, and as a particular case of interest the groups of elliptic curves. The following document presents the development of the elementary theory of elliptic curves and its use in the development of cryptosystems and implementation.

RESUMEN. Desde la introducción del concepto de llave publica en la criptografía, se ha desarrollado el potencial de utilizar el problema del logaritmo discreto, el cual se define como el problema de calcular logaritmos con respecto al generador en un grupo multiplicativo de enteros modulo primo. Las ideas propuestas para abordar dicho problema pueden ser extendidas a grupos arbitrarios, y como caso particular de interés los grupos de curvas elípticas. En el siguiente documento se presenta el desarrollo de la teoría elemental de las curvas elípticas y su uso en el desarrollo de criptosistemas e implementación.

1. Introducción

En tiempos modernos, el desarrollo de la tecnología ha permitido ampliar las formas de transmisión de información que ahora se han incorporado como parte fundamental del quehacer diario, ya sea en envío de un correo electrónico, mensajes de texto, pagos con tarjetas de crédito, etc. Todas estas actividades requieren un cierto nivel protección y seguridad de la información que se está comunicando, es en este punto donde la criptografía resulta ser útil para dicho propósito. En este documento se presenta una introducción de los conceptos y herramientas matemáticas básicas para el desarrollo teórico e implementación de criptosistemas basados en curvas elípticas. Para el desarrollo de la temática se comenzará presentando las ideas alrededor del concepto de llave pública y el problema de logaritmo discreto presentados por los matemático Diffie y Hellman [1], para luego abordar la idea del uso de grupos de puntos sobre una curva elíptica definida sobre un campo finito en un criptosistema de logaritmo discreto propuesta de forma independiente por N. Koblitz [2] y V. Miller [3]. La implementación de los sistemas de curvas elípticas sobre el grupo multiplicativo de un campo finito presenta como ventaja principal la ausencia de un algoritmo de tiempo sub - exponencial (como los de tipo "index-calculus") para encontrar logaritmos discretos en estos grupos. En consecuencia, al utilizar un grupo de curvas elípticas que es más pequeño en tamaño y manteniendo el mismo nivel de seguridad, resulta en tamaños de clave más pequeños, ahorro de

Fecha: August 15, 2023.
Palabras y frases clave. Criptografía, Llave pública, Curvas Elípticas.

ancho de banda e implementaciones más rápidas, características que son especialmente atractivas para aplicaciones de seguridad donde la potencia computacional y el espacio de almacenamiento es limitado, como pueden ser tarjetas inteligentes, computadores personales, dispositivos inalámbricos, etc.

1.0.1. *Justificación.* Por lo anterior expuesto es importante señalar los beneficios de la investigación en el área de la criptografía, ya que desempeña un papel fundamental en la sociedad actual debido a su impacto en múltiples aspectos de nuestra vida diaria. A medida que el mundo se vuelve cada vez más digitalizado y conectado, las aplicaciones de la criptografía se vuelven indispensables para garantizar la seguridad, integridad y privacidad de la información en diversos ámbitos, lo que contribuye de forma directa al desarrollo social, industrial y científico de un país. Lo que resulta acorde con los intereses académicos y vinculativos con la sociedad hondureña de la Universidad Nacional Autónoma de Honduras.

2. Antecedentes

2.1. **La criptografía moderna.** Se considera el nacimiento de la criptografía moderna con la publicación del artículo "Communication Theory of Secrecy Systems" por parte del matemático Claude Shannon [4] donde se comienza a tratar la criptografía desde un punto de vista matemático a través de lo que se denomina la teoría de la información. A partir de este punto la criptografía comienza a desarrollar con el único propósito de transmisión de información secreta, pero a finales del siglo XX el desarrollo de la informática comienza a dar lugar a nuevas aplicaciones de la criptografía como son los protocolos de autenticación, seguridad de transacciones electrónicas, verificación de integridad de datos, pruebas de conocimiento cero, etc. Todo el desarrollo comienza a requerir un trabajo conjunto con lo que se denominó criptoanálisis [5] (estudio de los sistemas criptográficos y sus debilidades) para en conjunto describir lo que es la criptología [6] (conjunto de técnicas para la transmisión de información de forma secreta), de aquí derivamos en uno de los conceptos fundamentales de la teoría criptográfica, la definición formal de un criptosistema [7].

Definición 2.1. Un criptosistema es una 5 - tupla $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ que satisface las siguientes condiciones:

- $\mathcal{P}$: es el conjunto finito posible de texto;
- $\mathcal{C}$: es el conjunto finito posible de texto cifrado;
- $\mathcal{K}$: es el conjunto finito de llaves posibles;
- Para cada $K \in \mathcal{K}$, existe una llave de encriptación $e_K \in \mathcal{E}$ y le corresponde una regla de desencriptación $d_K \in \mathcal{D}$. Cada $e_K : \mathcal{P} \to \mathcal{C}$ y $d_K : \mathcal{C} \to \mathcal{P}$ son funciones tales que $d_K(e_K(x)) = x$ para todo posible texto $x \in \mathcal{P}$.

De la definición se comienza a clasificar los criptosistemas en, sistemas de llave privada y sistemas de llave pública, los criptosistemas de llave privada son sistemas en los que las partes que intercambian información conocen las funciones e_K y d_k o la llave K de encriptación y en el casos de los sistemas de llave publica, cada parte que intercambia información tiene un par de funciones (e_K, d_k) de donde e_K es pública y donde d_K es la llave privada la cual depende directamente de e_K pero no es computacionalmente factible determinar d_K a partir de e_k, este es el esquema propuesto por los matemáticos Diffie y Hellman (1976)[1], donde las funciones e_K son la exponenciación en un grupo multiplicativo finito y determinar d_K requiere

resolver el problema del logaritmo discreto.
Una de las desventajas del modelo de llave publica es que se considera muy lento, los mismos autores proponen protocolos de intercambio de llaves junto con el las ideas de la firma digital para solventar dicha desventaja, que a pesar de las mismas desventajas, es un concepto ampliamente utilizado que an dado lugar a nuevos tipos de criptosistemas, por ejemplo el criptosistema de Massey - Omura [8], el criptosistema de ElGamal [9] y el criptosistema de Ciss - Cheikh - Sow [10].
En (1985) de forma independiente N. Koblitz [1] y V. Miller [3] proponen el uso de grupos de puntos en curvas elípticas definidas sobre campos finitos en criptosistemas de logaritmo discreto, que ha dado paso formas más seguras de aplicaciones ya conocidas de la criptografía y nuevas aplicaciones que han surgido con los recientes avances tecnológicos como son la seguridad en comunicación de vehículos autónomos, voto electrónico, blockchain, criptomonedas, reconocimiento óptico, etc. [11].

3. Conceptos Básicos

3.1. **Curvas Elípticas y leyes de grupo.**

3.1.1. *Curva Elíptica.*

Definición 3.1. Una curva elíptica E sobre un campo K, es una curva con una ecuación de la forma $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ para coeficientes apropiados $a_1, a_2, a_2, a_3, a_4, a_6 \in K$. Esta expresión es llamada la forma Weierstrass de E.

Será de nuestro interés, considerar los campos de números racionales $\mathbb{Q}$, números reales $\mathbb{R}$, números complejos $\mathbb{C}$ y el de los números enteros modulo p es decir $\mathbb{F}_q = \mathbb{Z}/p\mathbb{Z}$. También se debe mencionar que la forma definida anteriormente no es la forma más simple de una curva elíptica, siempre que la característica de K no sea 2 o 3, podemos simplificar la expresión completando el cuadrado con respecto a y y completando el cubo con respecto a x, esto lo podemos llevar a cabo considerando $y' = y + \left(\frac{a_1}{2}\right)x + \left(\frac{a_3}{2}\right)$ y $y' = x + \left(\frac{a_3}{3}\right)$, para concluir $(y')^2 = (x')^3 + A(x') + B$ forma a la cual se le llamaremos reducción de la forma de la forma de Weierstrass, la cual resulta mucho más sencilla de manejar en términos de computo.
Una característica importante que en general cumplen las curvas elípticas de la forma $y^2 = x^3 + Ax + B$ es que siempre son simétricas con respecto al eje x, es decir que se verifica que si (x, y) satisface la ecuación también $(x, -y)$.(Ver figura 1)
Uno de los elementos geométricos que será de nuestro interés es la recta tangente a la curva E, la cual puede ser determinada utilizando derivación implícita esto es $y' = \dfrac{3x^2 + A}{2y}$, expresión que nos permite establecer que $y' = \infty$ cuando $y = 0$ dado que $3x^2 + A$ no sea cero, esto puede ocurrir cuando $x^3 + Ax + B$ tiene raíces en común con su derivada $3x^2 + A$ lo que es equivalente a decir que $x^3 + Ax + B$ tiene doble raíz.

Definición 3.2. Si el polinomio $x^3 + Ax + B$ tiene raíces repetidas, decimos que la curva elíptica $y^2 = x^3 + Ax + B$ es singular. En caso contrario decimos que la curva elíptica es no singular. Una curva es singular si y solo si su discriminante $\Delta = -16(4A^3 + 27B^2) = 0$

El punto singular de la curva geométricamente ocurre cuando la curva se cruza consigo misma, a este tipo de singularidad la llamamos nodo y algebraicamente cuando el polinomio $x^3 + Ax + B$ tiene doble raíz.(Ver figura 2)

3.1.2. *Leyes de Grupo.* La propiedad fundamental de una curva elíptica y lo que la hace útil, es que si tenemos dos puntos en la curva, podemos construir un tercer punto que pertenece a la curva. De esta idea podemos establecer los pasos a considerar para determinar dicho punto:

(1) Suponemos $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$ dos puntos distintos sobre una curva elíptica $E : y^2 = x^3 + Ax + B$.
(2) Trazamos una recta que pasa por los puntos P_1 y P_2, asegurando que la recta L interseca E en un tercer punto Q. Suponemos que la recta L es de la forma $y = mx + b$ y que no es vertical. (Se considerará posteriormente dicha situación)
(3) La intersección entre L y E son las soluciones al sistema
$$\begin{cases} y = mx + b \\ y^2 = x^3 + Ax + B \end{cases}$$
lo que es equivalente a resolver la ecuación cubica
$$x^3 + (-m)x^2 + (A - 2mb)x + (B - b^2) = 0$$
sin embargo, conocemos dos raíces $x = x_1$ y $x = x_2$, por lo que se tiene garantizado el tercer punto Q.

Una vez considerados estos pasos para determinar un tercer punto, podemos utilizar la propiedad de simetría de la curva para determinar otro punto, esto es, si $P = (x, y)$ pertenece a la curva entonces el punto $-P = (x, -y)$ también pertenece a la curva.

Definición 3.3 (Ley de Grupo - 1)**.** Si P_1 y P_2 son dos puntos distintos sobre la curva elíptica $E : y^2 = x^3 + Ax + B$, sea $Q = (x', y')$ un tercer punto de intersección de E con la recta L que pasa por los puntos P_1 y P_2. Se define la suma $P_1 + P_2$ como el tercer punto $-Q = (x', -y')$.

Notar que la suma $P_1 + P_2$ no es la suma de componentes correspondientes entre los puntos P_1 y P_2. También se debe hacer notar que si consideramos dos puntos sobre $y^2 = x^3 + Ax + B$ tales que uno es reflexión vertical del otro, la recta no interseca la curva nuevamente, situación que será remediada considerando un punto sobre la curva al que llamamos punto en infinito en que denotaremos simplemente por ∞ y que se considera como punto de la recta vertical.

Ejemplo 3.4. Considerar los puntos $P_1 = (1, 2)$ y $P_2 = (3, 4)$ sobre la curva elíptica $y^2 = x^3 - 7x + 10$, para determinar la suma $P_1 + P_2$ y la suma $(P_1 + P_2) + P_2$. Siguiendo el esquema de construcción dado, tenemos que:

(1) La ecuación de la recta L tiene la forma $y = x + 1$.
(2) Lo que implica resolver
$$\begin{cases} y = x + 1 \\ y^2 = x^3 - 7x + 10 \end{cases} \implies x^3 - x^2 - 9x + 9 = 0 \implies (x - 1)(x - 3)(x + 3) = 0$$
obteniendo la coordenada $x = -3$ y por tanto el punto $Q = (-3, -2)$.
(3) De lo anterior concluimos $P_1 + P_2 = -Q = (-3, 2)$.

Ahora, para determinar $(P_1+P_2)+P_2$ seguimos un procedimiento similar al anterior:

(1) La ecuación de la recta L tiene la forma $y = \frac{1}{3}x + 3$.
(2) Lo que implica resolver

$$\begin{cases} y = \frac{1}{3}x + 3 \\ y^2 = x^3 - 7x + 10 \end{cases} \implies x^3 - \frac{1}{9}x^2 - 9x + 1 = 0 \implies \left(x - \frac{1}{9}\right)(x+3)(x-3) = 0$$

obteniendo la coordenada $x = \frac{1}{9}$ y por tanto el punto $Q = \left(\frac{1}{9}, \frac{82}{27}\right)$.

(3) De lo anterior concluimos que $(P_1 + P_2) + P_2 = \left(\frac{1}{9}, -\frac{82}{27}\right)$. (Ver figura 3)

Ahora se debe considerar definir el caso en que se desea calcular $P + P$, este casos se puede entender como la aproximación de P a P_1 es decir $P \to P_1$ de esta idea es que podemos entender a la recta L como la recta tangente a E en el punto P_1 y tomar el punto $-Q$ de reflexión del punto de intersección de L con E.

Definición 3.5 (Ley de Grupo - 2)**.** Si P es cualquier punto sobre la curva $E : y^2 = x^3 + Ax + B$, sea $Q = (x', y')$ el punto de intersección de E con la recta L tangente a E en el punto P. Definimos la suma $P + P$ como el punto $-Q = (x', -y')$.

Ejemplo 3.6. Considerar los puntos $P_1 = (1, 2)$ y $P_2 = (3, 4)$ sobre la curva elíptica $y^2 = x^3 - 7x + 10$, para determinar la suma $P_2 + P_2$ y la suma $(P_2 + P_2) + P_1$. Siguiendo la definición, procedemos con los siguientes pasos:

(1) Utilizando la diferenciación implícita determinamos $y' = \frac{3x^2 - 7}{2y}$ con lo que podemos determinar la ecuación de la recta tangente $L : y = \frac{5}{2}x - \frac{7}{2}$.
(2) Luego debemos resolver

$$\begin{cases} y = \frac{5}{2}x - \frac{7}{2} \\ y^2 = x^3 - 7x + 10 \end{cases} \implies x^3 - \frac{25}{4}x^2 + \frac{21}{2}x - \frac{9}{4} = 0 \implies \left(x - \frac{1}{4}\right)(x+3)(x-3) = 0$$

obteniendo la coordenada $x = \frac{1}{4}$ y por tanto el punto $Q = \left(\frac{1}{4}, -\frac{23}{8}\right)$.

(3) De lo anterior concluimos que $P_2 + P_2 = \left(\frac{1}{4}, \frac{23}{8}\right)$

Ahora, procedemos a determinar $(P_2 + P_2) + P_1$ siguiendo la ley de grupo - I:

(1) La ecuación de la recta L que pasa por los puntos $P_2 + P_2$ y P_1 tiene la forma $y = -\frac{7}{6}x + \frac{19}{6}$.
(2) Luego resolvemos

$$\begin{cases} y = -\frac{7}{6}x + \frac{19}{6} \\ y^2 = x^3 - 7x + 10 \end{cases} \implies x^3 - \frac{49}{36}x^2 + \frac{7}{18}x - \frac{1}{36} = 0 \implies \left(x - \frac{1}{9}\right)\left(x - \frac{1}{4}\right)(x-1) = 0$$

obteniendo la coordenada $x = \frac{1}{9}$ y por tanto el punto $Q = \left(\frac{1}{9}, \frac{82}{27}\right)$.

(3) De lo anterior concluimos que $(P_2 + P_2) + P_1 = \left(\frac{1}{9}, -\frac{82}{27}\right)$ (Ver figura 4).

Observación: Notar que $(P_1 + P_2) + P_2 = (P_2 + P_2) + P_1$.

Theorem 3.7 (Ley de Grupo)**.** *Si K es cualquier campo y E es una curva elíptica definida sobre K, entonces para cualesquiera puntos P, P_1, P_2, P_3 sobre E, se satisface lo siguiente:*

(1) $P_1 + P_2 = P_2 + P_1$ - *Ley Conmutativa.*
(2) $(P_1 + P_2) + P_3 = P_1 + (P_2 + P_3)$ - *Ley Asociativa.*
(3) $P + \infty = \infty + P = P$ - *El punto ∞ es identidad.*
(4) $P + (-P) = (-P) + P = \infty$ - *El punto inverso de P.*

El teorema enunciado condensa los razonamientos presentados anteriormente y ademas establece que bajo las operaciones definidas sobre E y el punto ∞ se define un grupo abeliano.

Theorem 3.8 (Ley de Grupo - Forma Explícita)**.** *Sea $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$ son puntos sobre la curva elíptica $E : y^2 = x^3 + Ax + B$. Entonces $P_1 + P_2 = (x_3, y_3)$ donde $x_3 = m^2 - x_1 - x_2$, $y_3 = -m(x_3 - x_1) - y_1$ y*

$$m = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & si \quad P_1 \neq P_2 \\ \dfrac{3x_1^2 + A}{2y_1} & si \quad P_1 = P_2 \end{cases}$$

si m infinito, entonces $P_1 + P_2 = \infty$.

3.1.3. *Curvas Elípticas Modulo p.* En esta sección, se presentará la teoría desarrollada utilizando curvas elípticas modulo p donde p es un número primo.
Consideraremos en el desarrollo de la sección a $p \geq 5$. En este caso se debe señalar que la curva $y^2 = x^3 + Ax + B \mod p$ es no singular cuando su discriminante $\Delta = -16(4A^3 - 27B^2) \mod p$ es distinto de cero.
En particular, se debe notar que una curva de esta forma siempre será singular modulo 2 y de forma más general, los primos p para los cuales la curva es singular mod p, son los primos los primos que dividen el discriminante Δ.

Ejemplo 3.9. Consideraremos los puntos $P_1 = (1, 3)$, $P_2 = (0, 2)$ sobre la curva $y^2 = x^3 + 4x + 4 \mod 5$ y determinaremos $P_1 + P_2$, $P_1 + P_1$.

(1) $P_1 + P_2$: Siguiendo el esquema desarrollado para este caso, tenemos que:

- $P_1 + P_2 = (x, y)$
- $x = m^2 - 1 - 0 = 0 \mod 5 = 0 \ : \ m = \dfrac{2-3}{0-1} = 1$
- $y = -(1)(0 - 1) - 3 = -2 \mod 5 = 3$
- $P_1 + P_2 = (0, 3)$

(2) $P_1 + P_1$: Siguiendo el esquema desarrollado para este caso, se tiene que:

- $P_1 + P_1 = (x, y)$
- $x = m^2 - 1 - 1 = 4 - 2 = 2 \mod 5 = 2 \ : \ m = \dfrac{3+4}{2 \cdot 3} \mod 5 = 2$
- $y = (-2)(2-1) - 3 = -5 \mod 5 = 0$
- $P_1 + P_1 = (2, 0)$.

Observación: Se pueden determinar los puntos $(0, 2),\ (0, 3),\ (1, 2),\ (1, 3),\ (2, 0),\ (4, 2)$, -$(4, 3)$ y ∞ (Ver figura 5).

Al determinar los puntos sobre la curva elíptica E modulo p, es posible notar que el número de puntos determinados es cercano a p, esta observación nos lleva a considerar el siguiente teorema:

Theorem 3.10 (Hasse)**.** *Sea E una curva elíptica no singular definida sobre un campo finito K con q elementos, entonces el número de puntos $N_q(E)$ en E cuyas componentes están en K, satisface $|N_q(E) - q - 1| \leq 2\sqrt{q}$.*

Podemos estimar la número de puntos que se deben esperar sobre una curva elíptica modulo p, considerando que para x existe p posibles valores y para y pueden ser 2, 1 o 0 dependiendo si, x es un valor cuadrado no cero, cero o no cuadrado respectivamente.
Considerando el hecho que si p (primo) $p \geq 3$ existen $\frac{p-1}{2}$ cuadrados no cero modulo p, de forma que el números de valores esperados para y dado un x particular es $\frac{1}{p}\left[2 \cdot \frac{p-1}{2} + 1 + 0 \cdot \frac{p-1}{2}\right] = 1$ y dado que son posibles p valores podemos esperar a lo más $2p$ junto con el punto ∞ sobre E, es decir $2p + 1$.
De lo anterior podemos establecer la desigualdad $1 \leq N_p(E) \leq 2p+1$ y reescribirla a la forma $|N_p(E) - p - 1| \leq p$. El teorema de Hasse mejora esta desigualdad estableciendo la $2\sqrt{p}$, una cota menor y mucho más cercano al valor que se obtienen de puntos sobre E.

3.1.4. *Orden de puntos.* El objetivo de esta sección es mostrar que el conjunto de puntos en una curva elíptica modulo p es un grupo abeliano finito bajo la operación de adición. Es requiere la siguiente definición:

Definición 3.11. Suponer la curva elíptica E definida sobre un campo K y un punto P sobre E.

(1) Para cualquier entero positivo k, se define el punto kP como la suma $\underbrace{P + P + \cdots + P}_{\text{k - elementos}}$ y definimos $(-k)P$ como el inverso aditivo $-kP$ junto con $0P = \infty$.

(2) El menor entero k para el cual $kP = \infty$ el llamado el orden de P y si tal valor no existe, decimos que P tiene orden infinito.

(3) Un punto de orden finito es llamado un punto de torsión y un punto con $mP = \infty$ es llamado un punto de m-torsión.

Ejemplo 3.12. Determinaremos el orden el punto $P = (1, 3)$ sobre la curva $E : y^2 = x^3 + 4x + 4 \mod 5$, comenzamos considerando $2P = P + P = (2, 0)$ punto determinado en ejemplo anterior, por lo que se tiene:

- $3P = 2P + P = (1, 2)$:
  Dado que $m = \dfrac{3-0}{1-2} = -3$ entonces $x = m^2 - 2 - 1 \mod 5$
  $= 6 \mod 5 = 1$ y $y = -(-3)(1-2) - 0 = -3 \mod 5 = 2$.

- $4P = 3P + P = \infty$:
  Dado que $x_1 = 1$ y $x_2 = 1$ podemos establecer que m es infinito, por lo que $4P = \infty$.

con lo que podemos concluir que el punto $P = (1, 3)$ tiene orden 4.

Observación: Podemos acelerar el proceso realizando el esquema $2P$, $4P$, $8P$, $\ldots, 2^j P$ junto con la diferencia, además en el caso de curvas elípticas sin incluir el modulo, podemos hacer uso de la diferencia de forma más inmediata, es decir si conocemos $P = (x, y)$ tenemos que $-P = (x, -y)$.

Ahora se enunciaran algunas de propiedades del orden de puntos sobre curvas elípticas:

Theorem 3.13. *Suponer la curva elíptica E y P un punto sobre E.*

(1) *Si P tiene orden finito y $mP = \infty$, entonces k divide m.*
(2) *Si $mP = \infty$ pero $\left(\frac{m}{q}\right) P \neq \infty$ para cualquier primo divisor q de m, entonces P tiene orden m.*
(3) *Si E es una curva elíptica modulo un primo p y N es el número de puntos en E modulo p, entonce $NP = \infty$, en particular el orden de P divide a N.*

Ejemplo 3.14. Mostraremos que el punto $P = (1, 3)$ tiene orden 15 en la curva elíptica $E : y^2 = x^3 + 4x + 4 \mod 13$, esto requiere utilizar la estrategia de calcular puntos con multiplicidad potencias de 2, podemos establecer
$2P = (12, 8)$, $4P = (6, 6)$, $8P = (0, 11)$, $16P = (1, 3)$, y verificar que el punto $15P = 16P - P = \infty$.

3.1.5. *Factorización con curvas elípticas.* En esta sección procederemos a desarrollar las ideas sobre el uso del algoritmo $p - 1$ de factorización de John Pollard [13] utilizando curvas elípticas, ideas originalmente propuesta por Hendrik Lenstra [14].

Consideramos una curva elíptica modulo N, donde N es un entero no primo, tal que el anillo $\mathbb{Z}/N\mathbb{Z}$ no es un campo. Sin embargo comenzamos con $E : y^2 = x^3 + Ax + B$ y $P = (a, b)$ punto sobre la curva $E \mod N$, lo que implica que $b^2 \equiv a^3 + Aa + B \mod N$ de forma que podemos aplicar las formulas para determinar $2P$, $3P$, $4P, \ldots$ que requieren la suma, resta, multiplicación y división por números que son primos relativos a N.

Ejemplo 3.15. Sea $E : y^2 = x^3 + 3x + 7 \mod 187$ y el punto $P = (38, 112)$ punto sobre E, calcularemos $2P$:
$x = m^2 - 38 - 38 = 10328 \mod 182 = 43 \ : \ m = \dfrac{3(38)^2 + 3}{2 \cdot 112} \mod 187 = 102$
$y = -(102)(43 - 38) - 112 \mod 187 = 126$ con lo que tenemos $2P = (43, 126)$.

Ahora procedemos a calcular $3P$ en forma similar a ejemplos anteriores de donde obtenemos $3P = (54, 105)$ y procedemos a calcular el punto $5P = 3P + 2P$ esto es: $m = \dfrac{105 - 126}{54 - 43} = \dfrac{-21}{11} \mod 187$, de donde debemos notar que no es posible calcular el recíproco de 11, dado que $MCD(187, 11) = 11$, por lo que no es posible calcular $5P$.
El resultado anterior nos revela la idea detrás del algoritmo de factorización con curvas elípticas, ya que el no poder calcular el punto $5P$ nos revela que 11 es divisor de 187 y nos permite determinar que $187 = 11 \cdot 17$.
Si consideramos la curva $E \mod 11$ y el punto $P = (38, 112) \equiv (5, 2) \mod 11$, podemos verificar que $5P = \infty$ en $E(\mathbb{F}_{11})$, es decir que al intentar calcular $5P \mod 11$ obtenemos el punto ∞, dado que en el proceso de calculo intentaremos realizar una división por cero. Pero en este contexto cero es el cero en $\mathbb{F}_{11}$ por lo que, realmente estamos determinado es el inverso multiplicativo ($\mod 11$) de algún entero que es divisible por 11.

Basados en las ideas presentadas en el ejemplo anterior junto con las ideas del algoritmo de Pollard se presenta el siguiente algoritmo de factorización utilizando curvas elípticas [14]:

Algorithm 1 Algoritmo de Lestrand para factorización por curvas elípticas

Require: N: Valor entero a factorizar.
1. Sea $P = (a, b)$ y $B \equiv b^2 - a^3 - A \cdot a(\mod N)$.
2. Sea E una curva elíptica $E : y^2 = x^3 + Ax + B$.

loop $j = 2, 3, 4, \ldots$ hasta $j <$ Alguna cota superior
- 3. Calcular $Q \equiv j \cdot P (\mod N)$ y $P = Q$
- **if** Calculo en (3) falla **then**
  - 4. Se determino $d > 1$ con $d \mid N$
- **end if**
- **if** $d < N$ **then**
  - 5. **return** d.
- **end if**
- **if** $d = N$ **then**
  - 6. Regresar a paso (1) y se selecciona una nueva curva E y punto P.
- **end if**

end loop incrementa $j \leftarrow j + 1$ e ir a paso (2)

Ejemplo 3.16. Consideraremos el entero $N = 6887$ y el punto escogido aleatoriamente $P = (1512, 3166)$ y el número $A = 14$, valores con los que realizamos el calculo $B \equiv 3166^2 - 1512^3 - 14 \cdot 1512 \mod 6887 = 19$.
Considerando los valores anteriores establecemos los valores para la curva $E : y^2 = x^3 + 14x + 19 \pmod{6887}$ la tiene por punto a P.
Ahora procedemos a calcular los puntos:

n	$n! \cdot P \mod 6887$
1	$1! \cdot P = (1512, 3166)$
2	$2! \cdot P = (3466, 2996)$
3	$3! \cdot P = (3067, 396)$
4	$4! \cdot P = (6507, 2654)$
5	$5! \cdot P = (2783, 6278)$
6	$6! \cdot P = (6141, 5581)$

Al considera $7!P$ debemos utilizar el resultado $6!P = (6141, 5581)$ y determinar $7P$ que para ello vamos establecer que $7P \equiv (p+2P)+4P \equiv (984, 589)+(203, 2038) \mod 6887$ dado que $P = (1512, 3166) \mod 6887$, $2P \equiv (3466, 2996) \mod 6887$ y $4P \equiv 2 \cdot 2P \equiv (208, 2038) \mod 6887$. Pero al calcular $7P$ se necesita determinar $\frac{589-2038}{203-984} \mod 6887$, lo que requiere determinar el inverso multiplicativo de $(-781)^{-1} \mod 6887$, pero se tiene que $\text{MCD}(-781, 6887) = 71$ lo que nos permite concluir que 71 es un divisor no trivial de 6887 y la factorización $6887 = 71 \cdot 91$.

3.2. **Encriptación con curva elíptica, residuo cuadrático.** El proceso de encriptación utilizando un punto de una curva elíptica no es un proceso trivial como lo sería en otro tipo de criptosistemas, por ello es necesario desarrollar algunos conceptos previos para lograr la encriptación deseada [15].

Definición 3.17. Si a es residuo modulo p, decimos que a es residuo cuadrático, si existe algún b talque $b^2 \equiv a \pmod{p}$ y si no existe tal b, decimos que a es un residuo no cuadrático.

Ejemplo 3.18. Considerando la definición anterior podemos considerar los siguientes residuos cuadráticos dado un primo p:

(1) modulo 5, el residuo cuadrático es $0, 1$ y 4, mientras que 2 y 3 no lo son.
(2) modulo 13, el residuo cuadrático es $0, 1, 4, 9, 3, 12$ y 10, mientras que $2, 5, 6, 7, 8$ y 11 no lo son.

A continuación consideramos un resultado de la teoría de números que puede resultar útil cuando determinamos residuos cuadráticos.

Theorem 3.19. *Existen* $\frac{p+1}{2}$ *residuos cuadráticos modulo* p *y son los valores* $0^2, 1^2, 2^2, \ldots, \left(\frac{p-1}{2}\right)^2$.

Considerando el segundo inciso en el ejemplo anterior vemos que existen $\frac{13+1}{2} = 7$ residuos cuadráticos, número que coincide con el calculado.

Otro concepto que nos resultara útil en la búsqueda de residuos cuadráticos es el símbolo de Legendre.

Definición 3.20. Si p es un primo impar y a un entero, el símbolo de Legendre denotado por $\left(\frac{a}{p}\right)_L$ y definido por $\left(\frac{a}{p}\right)_L = \begin{cases} -1 & \text{, Si } a \text{ no es residuo cuadráticos de } p \\ 0 & \text{, Si } p \text{ es divisor de } a \\ 1 & \text{, Si } a \text{ es residuo cuadrático de } p \end{cases}$

Ejemplo 3.21. Aquí presentamos un ejemplo sencillo del uso del la definición anterior:

(1) $\left(\frac{2}{7}\right)_L = 1$ dado que $2 \equiv 9 \mod 7$.

(2) $\left(\frac{7}{7}\right)_L = 0$ dado que $0 \equiv 49 \mod 7$.

(3) $\left(\frac{3}{7}\right)_L = -1$ dado que no es residuo cuadrático de 7.

Observación: La ecuación $x^2 \equiv a(\mod p)$ tiene exactamente $1 + \left(\frac{a}{p}\right)_L$ soluciones modulo p.

Theorem 3.22. *Si p es un primo impar, entonces para cualquier residuo clase a, se cumple que $\left(\frac{a}{p}\right)_L = a^{(p-1)/2}(\mod p)$.*

Ejemplo 3.23. Utilizando el teorema anterior verificaremos que $a = 17441$ es residuo cuadrático modulo $p = 239441$ y que $b = 135690$ no es residuo cuadrático modulo p:

(1) $a^{(239441-1)/2} = a^{119720} \equiv 1(\mod p)$, notando que $76197^2 \equiv 17441 \mod 239441$ y que $163244^2 \equiv 17441 \mod 239441$.

(2) $b^{(239441-1)/2} = b^{119720} \equiv -1(\mod p)$, no es residuo cuadrático modulo 239441 dado no es posible determinar x en $x^2 \equiv 135690(\mod 239441)$.

Theorem 3.24. *Dado cualquier primo impar p, el símbolo de Legendre modulo p es multiplicativo, es decir*

$$\left(\frac{a \cdot b}{p}\right)_L = \left(\frac{a}{p}\right)_L \cdot \left(\frac{b}{p}\right)_L$$

En particular, el producto de residuos no cuadráticos es un residuo cuadrático.

Theorem 3.25. *Si p es primo congruente a $3 \mod 4$ y a es residuo cuadrático modulo p, entonces $x = a^{(p+1)/4}$ tiene $x^2 \equiv a(\mod p)$*

Tomando los resultados presentados, proseguimos con el objetivo de realizar encriptación utilizando curvas elípticas. Se considerará la curva elíptica
$E : y^2 = x^3 + Ax + B \mod p$ donde es tal que $p \equiv 3(\mod 4)$.
Seguiremos el siguiente esquema para realizar la encriptación de un mensaje como parte de la coordenada x de un punto, para después buscar el resto de la coordenada de forma que x cumpla que $x^3 + Ax + B$ sea residuo cuadrático modulo p:

(1) Si p tiene $r + k + 1$ bits cuando escribimos en base 2, dividimos el mensaje en partes que contiene r bits.
(2) Para convertir un mensaje m de r bits, generamos una cadena previa que ira previos a m, esta es una cadena de $k + 1$ bits un cero seguido de k bits de la forma $0b_1b_2 \cdots b_km$.
(3) Luego buscamos entre las posibles elecciones de los k bits, hasta encontrar una solución y para $y^2 = x^3 + Ax + B(\mod p)$ y seleccionamos uno de los posibles valores de y arbitrariamente, realizando la encriptación utilizando $E \mod p$ y el punto (x, y).
(4) Para recuperar el mensaje m a partir del punto (x, y) donde $0 \leq x < p$, simplemente calculamos $x \mod 2^r$ y escribimos el resultado como una cadena de bits en base 2.

Ejemplo 3.26. Consideremos $m = 13 = 1101_2$ mensaje a ser encriptado con la curva elíptica $y^2 = x^3 + 11x + 17 \mod 307$ utilizando una longitud de $r = 4$ bits y un longitud de relleno de $k = 4$ bits.

(1) Notar que $p > 256 = 2^8$ por lo que p tiene 9 bits en base 2.
(2) Ahora procedemos a determinar una cadena $b_1b_2b_3b_4$ tal que $x = 0b_1b_2b_3b_41101_2$ sea residuo cuadrático modulo 307.
(3) Considerando $x = 000001101_2 = 13$ tenemos que $x^3 + 11x + 17 \equiv 208(\mod 307)$ de donde tenemos que no es residuo cuadrático modulo 307 dado que $208^{153} \equiv -1(\mod 307)$, pero considerando $x = 000011101_2 = 29$ se verifica que $x^3 + 11x + 17 \equiv 165(\mod 307)$ es residuo cuadrático dado que $165^{153} \equiv 1(\mod 307)$.
(4) Luego procedemos al calcular y, esto es $x^{(p+1)/2} \equiv 29^{77} \equiv 120(\mod 307)$ para concluir en el punto asociado a m es el punto $(29, 120)$ sobre E.
(5) Para recuperar m, simplemente consideramos la componente x del par ordenado y reducirlo a modulo 2^4, esto es obtener $13 \equiv 29 \mod 2^4$ el cual es el mensaje original.

3.3. **Encriptación de llave pública con curvas elípticas.** en esta sección presentaremos el uso del esquema de encriptación ElGamal aplicado al esquema de llave pública propuesto por Diffie-Hellman [9]:
Nota: El algoritmo considera a dos entes emisor y receptor, los cuales suponemos desean intercambiar información de forma segura.

Algorithm 2 Algoritmo ElGamal para esquema de llave pública Diffie-Hellman por curvas elípticas

▷ Generación de claves:

1. Elegir una curva elíptica E definida sobre un campo finito $\mathbb{F}_p$.
2. Elegir un punto base P en la curva E con un orden grande.
3. Elegir un entero d como llave privada, donde $1 < d <$ Orden de P.
4. Calcular el punto $Q = d \cdot P$.
5. La clave de llave pública será (E, p, P, Q), la cual se puede compartir para intercambiar los mensajes cifrados.

▷ Cifrado:

1. Convertir el mensaje m en un punto M en la curva E.
2. El remitente elige un valor temporal k aleatorio en el rango
$$1 < k < \text{ Orden de } P.$$
3. Calcular $C_1 = k \cdot P$.
4. Calcular $C_2 = M + k \cdot Q$, donde $+$ es la operación de suma en la curva elíptica.

El mensaje cifrado resultante es el par (C_1, C_2), el cual se envía al receptor.

▷ Descifrado:

1. Calcular $S = d \cdot C_1$.
2. Calcular $M = C_2 - S$, donde $-$ es la operación de resta en la curva elíptica.
3. Se interpreta M como el mensaje m enviado por el remitente.

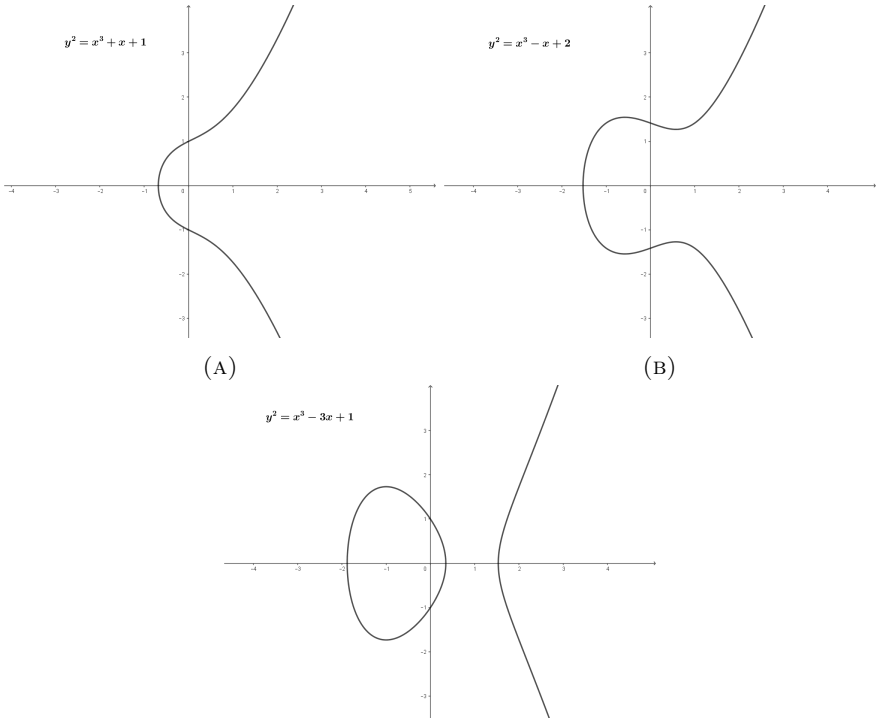
Ejemplo 3.27. Consideraremos la curva elíptica $E : y^2 = x^3 + 7x + 1$, $p = 44927$, $P = (7772, 14369)$ y $d = 22105$, para la generación de llaves, el mensaje a codificar $M = (14605, 29833)$ y descodificar. Asumimos la situación en la que se desea compartir un mensaje entre un emisor y un receptor utilizando el algoritmo presentado:

(1) Se determina $Q = d \cdot P = (39061, 4109)$, para establecer la llave pública (E, p, P, Q).
(2) Deseamos codificar el mensaje M, escogiendo un entero aleatorio k para este ejemplo $k = 23207$. Por lo que se calcula $C_1 = k \cdot M = (30566, 37885)$ y $C_2 = k \cdot Q + M = (35487, 8262) + P = (40194, 40273)$, lo que permite remitir el mensaje como (C_1, C_2) al receptor.
(3) El receptor al recibe el mensaje codificado (C_1, C_2), el cual procede a decodificar calculando $S = d \cdot C_1 = (35487, 36665)$ para determinar $M = C_2 - S = (14605, 29833)$ el mensaje original.

4. Conclusiones y trabajo a futuro

(1) El uso de curvas elípticas presenta similares beneficios que los criptosistemas tradicionales, pero con el beneficio de generar claves más cortas y fáciles de manejar, lo que permite el ahorro de recursos computacionales.
(2) Los criptosistemas basados en curvas elípticas presentan la dificultad de selección de parámetros, lo que puede comprometer la seguridad que provee, ya que dependen de algoritmos de selección aleatoria.
(3) De los aspectos anteriormente mencionados, se presenta la oportunidad de realizar trabajos para mejorar aspectos como la selección segura de parámetros, algoritmos más eficientes para realizar las operaciones sobre curvas elípticas.

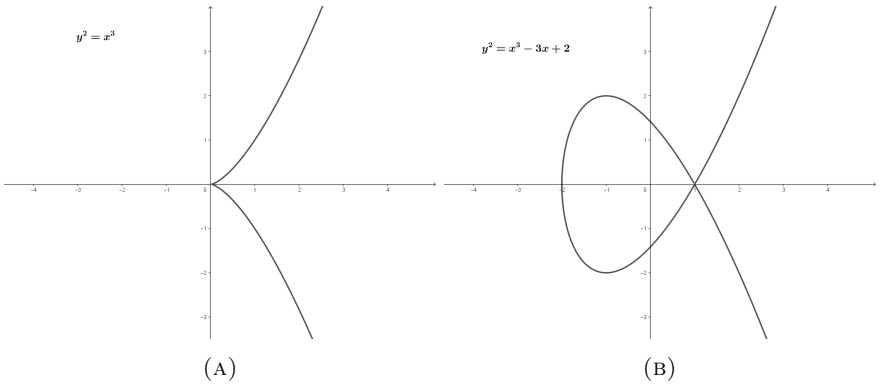
5. Figuras



(A) (B)

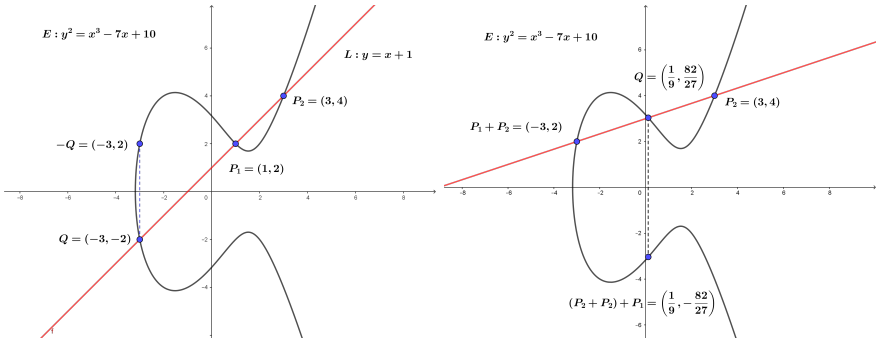
(C)

FIGURE 1. Ejemplos de Curvas Elípticas



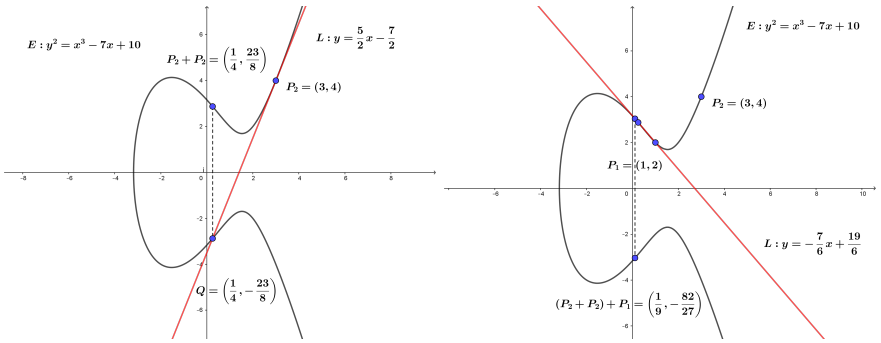
(A) (B)

FIGURE 2. Ejemplos de Curvas Elípticas - Singulares



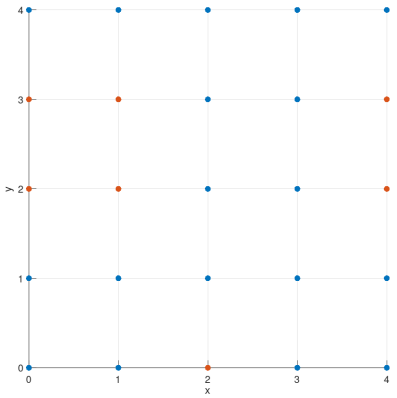
(A) (B)

FIGURE 3. Ejemplo - Ley de Grupo - I



(A) (B)

FIGURE 4. Ejemplo - Ley de Grupo - II



(A) Los elementos del grupo se resaltan en color rojo en la figura.

FIGURE 5. Ejemplo - $y^2 = x^3 + 4x + 4 \mod 5$

References

1. W. Diffie and M. Hellman, *New directions in cryptography, IEEE Transactions on Information Theory* Vol. 22 (1976) pp. 644–654.
2. N. Koblitz, *Primality of the number of points on an elliptic curve over a finite field* Pacific Journal of Mathematics, Vol. 131 (1988) pp. 157–165.
3. V. Miller, *Uses of elliptic curves in cryptography*, Advances in Cryptology—CRYPTO '85, Lecture Notes in Computer Science, Springer-Verlag, 218 (1986) pp. 417–426.
4. Claude E. Shannon, *Communication Theory of Secrecy Systems*, Bell System Technical Journal, vol.28-4, page 656–715, Oct. 1949.
5. William f. Friedman, *Army Extension Courses, Sub Course: Military Cryptanalysis Part I, Monoalphabetic Substitution System, Lesson Assignment Sheets*, 1936-1937
6. Wade Trappe, Lawrence C. Washington, *Introduction to cryptography with coding theory*, Prentice Hall 2002.
7. Douglas R. Stinson, *Cryptography: Theory and practice*, Discrete Mathematics and Its Applications, Chapman and Hall, 2005.
8. J. L. Massey y J. K. Omura., *Method and apparatus for maintaining the privacy of digital messages conveyed by public transmition* [en línea]. U.S. Patent #4,567,600, 28 de enero de 1986. Disponible de World Wide Web: *www.google.com/patents/US4567600.pdf*
9. T. ElGamal, *A public key cryptosystem and a signature scheme based on discrete logarithms*, IEEE Transactions on Information Theory 31, 469-472 (1985)
10. A.A. Ciss, A.Y. Cheikh y D. Sow, *A Factoring and Discrete Logarithm based Cryptosystem*, International Journal of Contemporary Mathematical Sciences 8(11), 511 - 517 (2013)
11. Shamsher Ullah, Jiangbin Zheng, Nizamud Din, Muhammad Tanveer Hussain, Farhan Ullah, Mahwish Yousaf, *Elliptic Curve Cryptography; Applications, challenges, recent advances, and future trends: A comprehensive survey*, Computer Science Review, 47 (2023) 100530.
12. N. Koblitz, A Course in Number Theory and Crytography, Springer - Verlang, 2nd edition, 1994.
13. Pollard, J. (1974). Theorems on factorization and primality testing. Mathematical Proceedings of the Cambridge Philosophical Society, 76(3), 521-528. doi:10.1017/S0305004100049252
14. H. W. Lenstra, Jr. Factoring integers with elliptic curves. Ann. Math. 126 (1987), 649- 673.
15. J.H. Silverman, Jill Pipher, Jeffrey Hoffstein, *An Introduction to Mathematical Cryptography*, Springer Science + Business Media, LLC, part of Springer Nature 2008.

Maestría en Matemática, Universidad Nacional Autónoma de Honduras
Dirección de correo electrónico: carlos.urrutia@unah.edu.hn