



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

**INSTITUTO TECNOLÓGICO DE CIUDAD MADERO**  
**DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**  
**MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**



"POR MI PATRIA Y POR MI BIEN"

**TESIS**

**OPTIMIZACIÓN DEL PROBLEMA DE SELECCIÓN DE CARTERA DE  
PROYECTOS CON PARÁMETROS DIFUSOS TRIANGULARES**

Que para obtener el Grado de  
**Maestro en Ciencias de la Computación**

Presenta

**Ing. Luis Mario Velasco Ocejo**

**G14071407**

**1006599**

Director de Tesis

**Dr. Héctor Joaquín Fraire Huacuja**

**123070**

Co-director de tesis

**Dr. Fausto Antonio Balderas Jaramillo**

Ciudad Madero, Tamaulipas

Noviembre 2021



Cd. Madero, Tam. **20 de octubre de 2021**

**OFICIO No. : U.149/21**  
**ASUNTO: AUTORIZACIÓN DE IMPRESIÓN DE TESIS**

**C. LUIS MARIO VELASCO OCEJO**  
**No. DE CONTROL G14071407**  
**P R E S E N T E**

Me es grato comunicarle que después de la revisión realizada por el Jurado designado para su Examen de Grado de Maestría en Ciencias de la Computación, se acordó autorizar la impresión de su tesis titulada:

**“OPTIMIZACIÓN DEL PROBLEMA DE SELECCIÓN DE CARTERA DE PROYECTOS CON PARÁMETROS DIFUSOS TRIANGULARES”**

El Jurado está integrado por los siguientes catedráticos:

PRESIDENTA:	DRA. CLAUDIA GUADALUPE GÓMEZ SANTILLÁN
SECRETARIO:	DR. JUAN FRAUSTO SOLÍS
VOCAL:	DR. HÉCTOR JOAQUÍN FRAIRE HUACUJA
SUPLENTE:	DR. FAUSTO ANTONIO BALDERAS JARAMILLO
DIRECTORA DE TESIS:	DR. HÉCTOR JOAQUÍN FRAIRE HUACUJA
CO-DIRECTOR:	DR. FAUSTO ANTONIO BALDERAS JARAMILLO

Es muy satisfactorio para la División de Estudios de Posgrado e Investigación compartir con usted el logro de esta meta. Espero que continúe con éxito su desarrollo profesional y dedique su experiencia e inteligencia en beneficio de México.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica*

*"Por mi patria y por mi bien"*

**MARCO ANTONIO CORONEL GARCÍA**  
**JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**



c.c.p.- Archivo  
MACG 'jar'



# Declaración de originalidad

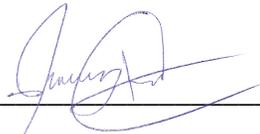
---

Declaro y prometo que este documento de tesis es producto de mi trabajo original y que no infringe los derechos de terceros, tales como derechos de publicación, derechos de autor, patentes y similares.

Además, declaro que en las citas textuales que he incluido (las cuales aparecen entre comillas) y en los resúmenes que he realizado de publicaciones ajenas, indico explícitamente los datos de los autores y las publicaciones.

Además, en caso de infracción de los derechos de terceros derivados de este documento de tesis, acepto la responsabilidad de la infracción y relevo de ésta a mi director y codirectores de tesis, así como al Tecnológico Nacional de México campus Ciudad Madero y sus autoridades.

4 de Noviembre de 2021, Cd. Madero, Tamps.



---

Ing. Luis Mario Velasco Ocejo

# Agradecimientos

---

Quiero agradecer a mis padres, Hugo y Paty por siempre estar conmigo, dándome la confianza y el apoyo que necesito. A mi hermano Hugo, le agradezco por ser mi mentor a lo largo de toda mi formación académica y, desde luego, también durante esta etapa relacionada a la investigación. A Robin, Juan y el resto de mi familia les agradezco por su apoyo y compañía en las buenas y en las malas, por las charlas técnicas, por los proyectos caseros y todos los momentos que hemos compartido. Este logro también es suyo.

A Bacho, mi viejito, que no pudo llegar a compartir este momento conmigo pero que me demostró lo que son el amor y cariño incondicionales. Que no importa lo que suceda y por lo que estemos pasando, cada día es una nueva historia, en donde habrá cosas por descubrir, cosas por las que reír y momentos que disfrutar. A donde se encuentre, todo mi agradecimiento.

Agradezco a mi director de tesis, el Doctor Héctor Joaquín Fraire Huacuja por brindarme la oportunidad de formar parte de su grupo de trabajo. Sus consejos, motivación y experiencia han sido fundamentales para la realización de este trabajo.

A los integrantes de mi comité tutorial, les agradezco por su retroalimentación y consejos. Han sido de gran ayuda durante esta etapa de mi formación y para la realización de este trabajo.

Por último, quiero agradecer a la vida por permitirme llegar a este momento habiendo recorrido un camino colmado de experiencias y aprendizaje.

# Resumen

---

El problema de selección de cartera de proyectos, al igual que todas aquellas situaciones que involucran la inversión de recursos, representa un proceso de vital importancia para las entidades organizacionales, ya que de la selección de una cartera apropiada dependerá la posible obtención de beneficios. Su resolución implica tomar en consideración distintos objetivos (a menudo en conflicto) y restricciones impuestas por la entidad organizacional, lo que permite abordarlo como un problema de optimización. Sin embargo, como suele ocurrir en diversas situaciones del mundo real, la información disponible acerca de dichos elementos puede verse afectada por un factor que puede provenir de distintas fuentes que van desde una pobre especificación de elementos, hasta factores externos que provocan la variabilidad de presupuestos y costos a través del tiempo: la incertidumbre.

Bajo este contexto, aquella entidad encargada de resolver el problema se verá forzada a trabajar con información que no describe de forma precisa los presupuestos, costos y beneficios que representan la inversión de recursos en una determinada cartera de proyectos. Para ello, dicha información deberá ser adecuadamente manipulada y modelada empleando alguna estrategia que permita su incorporación a modelos de optimización matemática. En la literatura, la incertidumbre presente en este problema ha sido modelada empleando una gran variedad de estrategias como intervalos, enfoques probabilísticos, conjuntos difusos, entre otros. La estrategia utilizada dependerá de las características de la incertidumbre.

Este trabajo de tesis aborda la incertidumbre presente en los parámetros (presupuestos, costos y beneficios) del problema de selección de cartera de proyectos empleando como estrategia de modelado números difusos triangulares, una estrategia escasamente utilizada de acuerdo a la revisión de trabajos relacionados. Dicha estrategia es utilizada para proponer un nuevo modelo matemático del problema. Además, se proponen dos algoritmos de solución basados en la metodología del algoritmo genético de no dominancia *Nondominated Sorting Genetic Algorithm II* (NSGA-II, por sus siglas en inglés): Tri-NSGA-II-CD y Tri-NSGA-II-SSD. Uno de ellos incorpora el estimador de densidad Desviación de Dispersión Espacial (*Spatial Spread Deviation*, SSD por sus siglas en inglés) que permite mejorar la distribución de soluciones en los frentes de Pareto.

Para evaluar el desempeño de los algoritmos propuestos, se realizaron una serie de experimentos sobre conjunto de instancias de distintos tamaños. Para dicha evaluación se usaron dos métricas clásicas: hipervolumen y dispersión generalizada. Las conclusiones de la evaluación experimental fueron validadas con la aplicación de dos pruebas no paramétricas de hipótesis: Wilcoxon y Friedman. Los resultados de la evaluación de dichas pruebas permiten observar el impacto de la incorporación del estimador de densidad SSD.

# Tabla de contenido

<b>CAPÍTULO 1: INTRODUCCIÓN .....</b>	<b>1</b>
1.1. MOTIVACIÓN .....	1
1.2. ANTECEDENTES .....	2
1.3. JUSTIFICACIÓN.....	3
1.4. OBJETIVOS .....	4
1.4.1 <i>Objetivo general</i> .....	4
1.4.2 <i>Objetivos específicos</i> .....	4
<b>CAPÍTULO 2: MARCO TEÓRICO .....</b>	<b>5</b>
2.1 PROBLEMAS DE OPTIMIZACIÓN MULTIOBJETIVO.....	5
2.2 DOMINANCIAS.....	6
2.3 OPTIMALIDAD DE PARETO.....	7
2.4 PROBLEMA DE SELECCIÓN DE CARTERA DE PROYECTOS .....	7
2.5 EL MODELADO DE LA INCERTIDUMBRE .....	9
2.6 ELEMENTOS DE LA TEORÍA DE NÚMEROS DIFUSOS .....	10
2.6.1 <i>Conjuntos difusos</i> .....	11
2.6.2 <i>Números difusos generalizados</i> .....	14
2.6.3 <i>Números difusos trapezoidales</i> .....	15
2.6.4 <i>Números difusos triangulares</i> .....	17
2.7 ARITMÉTICA DE NÚMEROS DIFUSOS.....	18
2.7.1 <i>Operaciones aritméticas entre números trapezoidales</i> .....	19
2.7.2 <i>Operaciones aritméticas entre números triangulares</i> .....	19
2.8 COMPARACIÓN DE NÚMEROS DIFUSOS .....	20
2.8.1 <i>Graded Mean Integration</i> .....	20
2.8.2 <i>Relación de orden entre números difusos</i> .....	21
2.8.3 <i>Dominancia entre vectores difusos</i> .....	21
2.9 METAHEURÍSTICAS POBLACIONALES .....	22
2.9.1 <i>Algoritmos genéticos</i> .....	22
2.10 PRUEBAS NO PARAMÉTRICAS .....	23
2.10.1 <i>Prueba de rangos con signo de Wilcoxon</i> .....	24
2.10.2 <i>Prueba de Friedman</i> .....	26
<b>CAPÍTULO 3: ESTADO DEL ARTE.....</b>	<b>28</b>
3.1 SOLUCIÓN DE PROBLEMAS DE CARTERA DE PROYECTOS PÚBLICOS A PARTIR DE INFORMACIÓN DEL RANKING DE PRIORIDADES .....	28
3.2 ARTIFICIAL BEE COLONY ALGORITHM BASED ON NOVEL MECHANISM FOR FUZZY PORTFOLIO SELECTION .....	29
3.3 FUZZY PORTFOLIO OPTIMIZATION MODEL UNDER REAL CONSTRAINTS .....	29
3.4 A NEW BI-OBJECTIVE FUZZY PORTFOLIO SELECTION MODEL AND ITS SOLUTION THROUGH EVOLUTIONARY ALGORITHMS.....	30
3.5 FUZZY PORTFOLIO OPTIMIZATION MODEL WITH FUZZY NUMBERS.....	30
3.6 MODELING AND OPTIMIZING THE MULTI-OBJECTIVE PORTFOLIO OPTIMIZATION PROBLEM WITH TRAPEZOIDAL FUZZY PARAMETERS .....	30

3.7	MODELANDO LA IMPRECISIÓN DEL PROBLEMA DE CARTERA DE PROYECTOS CON FILOSOFÍA GRIS	31
-----	--	----

**CAPÍTULO 4: MODELANDO EL PROBLEMA DE SELECCIÓN DE CARTERA DE PROYECTOS CON PARÁMETROS TRIANGULARES..... 34**

4.1	MODELO MATEMÁTICO.....	34
4.2	REPRESENTACIÓN DE SOLUCIONES Y VALIDACIÓN DE FACTIBILIDAD.....	35

**CAPÍTULO 5: ALGORITMO PROPUESTO: TRI-NSGA-II..... 38**

5.1	OPERADOR DE CRUZA: CRUZA UNIFORME.....	39
5.2	OPERADOR DE MUTACIÓN: <i>ONE BIT FLIP</i> .....	40
5.3	GENERACIÓN DE LA POBLACIÓN INICIAL .....	40
5.4	ORDENAMIENTO DE LA POBLACIÓN .....	40
5.4.1	<i>Non-dominated Sorting</i> .....	41
5.4.2	<i>Crowding Distance Sorting</i> .....	41
5.4.3	<i>Spatial Spread Deviation Sorting</i> .....	42

**CAPÍTULO 6: RESULTADOS EXPERIMENTALES ..... 43**

6.1	INSTANCIAS DEL PROBLEMA.....	43
6.1.1	<i>Conversión de instancias</i> .....	45
6.2	ALGORITMO EXHAUSTIVO .....	46
6.3	CONDICIONES DE EXPERIMENTACIÓN .....	48
6.3.1	<i>Hardware y Software</i> .....	48
6.3.2	<i>Configuración de parámetros de algoritmos</i> .....	48
6.4	VALIDACIÓN DE FUNCIONAMIENTO .....	48
6.5	EVALUACIÓN DE DESEMPEÑO .....	50
6.5.1	<i>Generación de frentes de referencia</i> .....	50
6.5.2	<i>Resultados de evaluación</i> .....	51

**CAPÍTULO 7: CONCLUSIONES Y TRABAJOS FUTUROS ..... 53**

7.1	CONTRIBUCIONES.....	53
7.2	PRODUCTOS CIENTÍFICOS .....	54
7.3	TRABAJOS A FUTURO.....	54

**REFERENCIAS ..... 55**

# Capítulo 1

## Introducción

El presente trabajo de tesis representa una aportación a las estrategias de modelado y resolución de la variante multiobjetivo del problema de selección de cartera de proyectos. En él se aborda el problema desde un enfoque centrado en modelar la imprecisión presente en los datos que describen una instancia particular del problema, tales como presupuestos, costos y beneficios. Como estrategia de modelado, se introduce una metodología basada en números difusos de tipo triangular. Además, se emplean métodos que permiten su resolución mediante algoritmos genéticos basados en la relación de no dominancia.

En este capítulo se describe de manera general el trabajo de tesis, enunciando sus objetivos, alcances, limitaciones y las razones que motivaron y justificaron su realización.

### 1.1. Motivación

En el entorno organizacional, dada la naturaleza cambiante de los mercados y el constante proceso de innovación tecnológica de la industria, es común que las organizaciones de cualquier rubro se enfrenten a problemas de decisión en los que tomar una buena o mala decisión influirá directamente sobre los recursos de los que dispone y el curso de acciones a tomar a corto, mediano o largo plazo. El proceso de toma de decisiones recae en una entidad conocida como tomador de decisiones o *Decision Maker* [1] (DM, por sus siglas en inglés) y deberá estar directamente influenciado por los objetivos y requisitos organizacionales.

El problema de selección de cartera de proyectos es un problema de optimización clásico que suele presentarse con frecuencia en organizaciones de cualquier índole debido a su necesidad de invertir en proyectos, ya sea de forma simultánea o consecutiva, con el propósito de tratar de garantizar su estabilidad o generarle oportunidades de crecimiento. El

problema se presenta cuando el costo de inversión del conjunto de proyectos supera el presupuesto de inversión del que la organización dispone, por lo que el DM deberá seleccionar aquel subconjunto de proyectos que representen el mayor beneficio a la organización empleando el presupuesto disponible. Dicho proceso de toma de decisiones deberá estar siempre precedido por una labor de análisis de los elementos que interactúan en el problema (proyectos y presupuestos).

Dado que el trabajo del DM está directamente ligado a la información disponible, puede resultarle especialmente complejo seleccionar una alternativa cuando dicha información presenta incertidumbre o imprecisión [1]. Las dificultades que pueden presentarse pueden variar de acuerdo a las características de la información, algunas de ellas son:

- La información se expresa de manera cualitativa (vaguedad): dificulta su modelado para incorporarlo a métodos de optimización.
- La información es imprecisa o incierta: no se conocen con exactitud datos del problema debido a la falta de información o a la incertidumbre presente al estimar estados futuros de los proyectos.

Actualmente, existen diversas estrategias que permiten modelar la presencia de incertidumbre en el problema de selección de cartera de proyectos, facilitando así su incorporación a diversos métodos de resolución. Entre ellas destacan el uso de enfoques probabilísticos, matemática de intervalos y números difusos. Dichas estrategias permiten modelar algunas de las formas en las que suele presentarse la incertidumbre.

La principal motivación de este trabajo es contribuir al conjunto de estrategias que permiten modelar y resolver el problema de selección de cartera de proyectos considerando la incertidumbre. Por ello, se enfoca en el uso de números difusos triangulares, estrategia útil para modelar algunos tipos de incertidumbre que se presentan con frecuencia en situaciones del mundo real.

## **1.2. Antecedentes**

Este trabajo de tesis forma parte de un conjunto de trabajos enfocados a la optimización del problema de selección de cartera de proyectos realizados en el Instituto Tecnológico de Ciudad Madero (ITCM). En dichos trabajos, se han propuesto diversas estrategias y métodos que permiten abordar el problema desde distintas perspectivas y en sus distintas variantes (*e.g.*: con múltiples objetivos, dinámicas, con incorporación de preferencias o con conocimiento imperfecto).

Una de las variantes del problema más estudiadas en tiempo recientes en el grupo de trabajo es aquella enfocada en modelar la incertidumbre presente en los parámetros del problema (costos, presupuestos y beneficios) mediante distintas estrategias y proponer métodos que permiten su incorporación a estrategias de resolución basadas en

metaheurísticas clásicas. Algunos autores han optado, por ejemplo: por usar conjuntos difusos para representar estados lingüísticos, intervalos para representar casos en los que los costos se ubican dentro de cierto rango [2] o números difusos trapezoidales [3].

Este trabajo parte de las bases sentadas por los trabajos que le anteceden para, de forma similar, proponer una estrategia más de modelado para tratar la incertidumbre de los parámetros del problema: los números difusos triangulares. Así mismo, propone su incorporación a dos algoritmos genéticos basados en relaciones de no dominancia. Características que le permiten enriquecer el conjunto de trabajos existentes y que constituyen una aportación al campo de investigación al tratar una forma más en la que puede presentarse la incertidumbre.

### **1.3. Justificación**

La presencia de la incertidumbre es un fenómeno ineludible en muchas de las situaciones que se presentan en el mundo real [1] y desde luego, los problemas de optimización como el de la selección de cartera de proyectos no son la excepción. Una representación adecuada de la incertidumbre permitirá la correcta consideración y evaluación de límites y restricciones durante el proceso de resolución. Por ello, modelar la incertidumbre representa una tarea especialmente importante, sobre todo en aquellos problemas que comprometen o involucran de alguna manera los recursos de los que dispone una entidad organizacional.

En diversos trabajos de investigación se han propuesto una gran variedad de estrategias que permiten modelar la incertidumbre presente en la formulación de varios problemas de optimización. Sin embargo, la incertidumbre no siempre proviene de la misma fuente o se presenta de la misma manera [1] [4]. Por ello, es probable que alguna de las estrategias existentes permita una mejor representación de determinado tipo de incertidumbre [5] o al menos una representación que resulte más intuitiva o cercana a la interpretación humana de los fenómenos inciertos. Por ejemplo, algunos autores [6] [7] han optado por aplicar una de las estrategias más ampliamente utilizadas para modelar la incertidumbre en distintos campos de la ciencia: la teoría de la probabilidad. Sin embargo, a pesar de su eficacia, no es posible modelar todas las distintas manifestaciones de la incertidumbre empleando únicamente la teoría de la probabilidad [8] (por ejemplo, la incertidumbre causada por el uso de expresiones en lenguaje natural como “muy rápido”). Por lo que, de acuerdo a las características del problema, deberán seleccionarse aquellas estrategias que permitan modelar formas específicas de la incertidumbre.

Para el caso particular del problema que se aborda en este trabajo se han propuesto una gran cantidad de estrategias de modelado de incertidumbre, algunas basados en la lógica difusa [9] [10], matemática de intervalos [11] o números difusos [3]; cada una de ellas adecuada para tratar distintas formas de incertidumbre. No obstante, de acuerdo con la revisión de trabajos relacionados, una de las estrategias menos exploradas es aquella cuyo modelado se basa en números difusos triangulares.

En esta tesis, se propone una estrategia que permite modelar e incorporar a algoritmos de solución la incertidumbre del problema de cartera de proyectos mediante números difusos de tipo triangular, estrategia comúnmente utilizada para modelar expresiones en lenguaje natural del tipo “*cercano a*” [5]. Generando así una aportación al catálogo de estrategias de modelado de incertidumbre del problema, lo que permite a la entidad encargada del análisis del mismo contar con una mayor cantidad opciones para elegir la más apropiada de acuerdo a la forma en que se presenta la incertidumbre.

## **1.4. Objetivos**

### **1.4.1 Objetivo general**

Diseñar e implementar un algoritmo NSGA-II para el problema de selección de cartera de proyectos con parámetros difusos del tipo triangular.

### **1.4.2 Objetivos específicos**

- 1) Revisar el estado del arte de la resolución del problema de selección de cartera de proyectos con parámetros difusos.
- 2) Revisar el marco teórico.
- 3) Diseñar un nuevo modelo del problema de optimización multiobjetivo de la selección de cartera de proyectos con parámetros difusos triangulares.
- 4) Diseñar e implementar un algoritmo exhaustivo para resolver el problema de selección de cartera de proyectos sin considerar regiones ni preferencias.
- 5) Diseñar e implementar un algoritmo NSGA-II con parámetros difusos triangulares para resolver el problema de selección de cartera de proyectos.

# Capítulo 2

## Marco teórico

Este capítulo concentra la definición de conceptos y estrategias que permiten abordar los problemas de optimización multiobjetivo bajo el contexto del modelado de la incertidumbre. Además, se introducen algunos de los conceptos básicos empleados para modelarla mediante números difusos de tipo triangular.

### 2.1 Problemas de optimización multiobjetivo

Un problema de optimización puede definirse en pocas palabras como la búsqueda de un máximo o mínimo (óptimo) para una determinada función. Función cuyos valores de variables pueden estar acotados a un determinado sector del espacio de búsqueda mediante el establecimiento de restricciones [12].

Los problemas de optimización nacen de la necesidad de adaptar sistemas, por lo regular industriales, de acuerdo a los objetivos de los usuarios (minimizar el consumo de energía, maximizar la producción de un producto, reducir gastos). Sin embargo, al modelar un problema de la vida real, por lo regular se desea satisfacer más de un objetivo (*e.g.*: un sistema eficiente y que además tenga un reducido consumo de energía). A este tipo de problemas de optimización se les conoce como problemas de optimización multiobjetivo. Un problema de optimización multiobjetivo puede modelarse matemáticamente de manera general como [13]:

$$\begin{aligned} & \text{Maximizar / Minimizar } f_m(x), & m = 1,2,3, \dots, M; \\ & \text{sujeto a: } & g_j(x) \geq 0, & j = 1,2,3, \dots, J; \\ & & h_k(x) = 0, & k = 1,2,3, \dots, K; \\ & & x_i^{(L)} \leq x_i \leq x_i^{(U)}, & i = 1,2,3, \dots, n; \end{aligned}$$

En donde existen  $M$  funciones objetivo  $f(x) = (f_1(x), f_2(x), \dots, f_M(x))$  que pueden ser maximización o de minimización. Una solución  $x$  es un vector de  $n$  variables de decisión  $x = (x_1, x_2, \dots, x_n)$ . Asociados al problema se encuentran  $J$  restricciones de desigualdad y  $K$  restricciones de igualdad dadas por  $g_j(x)$  y  $h_k(x)$  respectivamente. El último conjunto de restricciones acota los valores de las variables de decisión a un espacio de búsqueda entre los límites inferior y superior dados por  $x_i^{(L)}$  y  $x_i^{(U)}$  respectivamente. A dicho espacio se le conoce como espacio de decisión.

En la resolución de problemas de optimización multiobjetivo, al igual que en los problemas mono objetivo, se pretende alcanzar un escenario ideal: encontrar una única solución que lo resuelva (solución óptima). Sin embargo, en la mayoría de los problemas multiobjetivo los objetivos suelen ser contradictorios entre sí (el decremento de un objetivo significa el incremento de otro o viceversa). Por lo tanto, la meta radica en minimizar o maximizar el valor de los distintos objetivos “de la mejor forma posible”.

A pesar de que los problemas multiobjetivo carecen de una única solución que permita optimizar todos los objetivos (solución óptima), es posible un encontrar un conjunto de soluciones igualmente buenas o no inferiores entre sí conocido como el conjunto de soluciones óptimas de Pareto [14] o frente de Pareto.

## 2.2 Dominancia

Al resolver un problema de optimización multiobjetivo, sólo un pequeño subconjunto de posibles soluciones es de interés: el conjunto de soluciones que conforma el frente de Pareto. Para distinguir aquellas soluciones que conforman el frente de Pareto de todas las demás del espacio de soluciones es necesario hacer uso del concepto de dominancia.

La dominancia es un concepto que permite comparar soluciones de acuerdo a los valores de sus objetivos con el propósito de establecer una relación de superioridad o inferioridad entre ellas.

Se dice que una solución  $x^*$  domina (o supera) a una solución  $x$  si se cumplen las siguientes condiciones [15]:

- 1)  $f_i(x^*) \geq f_i(x)$  para todo  $i \in [1, k]$ .
- 2)  $f_j(x^*) > f_j(x)$  para al menos un valor de  $j \in [1, k]$ .

Es decir, que  $x^*$  es al menos tan bueno como  $x$  para los valores de todas las funciones objetivo y que es mejor que  $x$  en el valor de al menos una función objetivo. Para indicar que la solución  $x^*$  domina a la solución  $x$  se utiliza la siguiente notación:

$$x^* > x$$

En el caso que únicamente se cumpla la primera de las condiciones, se dice que  $x^*$  domina débilmente a  $x$  y se indica mediante la siguiente notación:

$$x^* \succcurlyeq x$$

Por otro lado, si no existe solución  $x$  que domine a la solución  $x^*$  se dice que  $x^*$  es una solución no dominada, no inferior o admisible.

## 2.3 Optimalidad de Pareto

Mediante el operador de dominancia es posible comparar todos elementos que conforman el espacio de soluciones de determinado problema. Aquellas soluciones que luego de dicha comparación, no sean dominadas por ninguna otra, se consideran soluciones óptimas del frente de Pareto. En el sentido de optimalidad del frente de Pareto, si  $x^*$  es una solución a un problema multiobjetivo:

$$x^* \text{ es óptima} \Leftrightarrow \nexists x: (f_i(x) \geq f_i(x^*) \forall i \in [1, k] \text{ y } f_j(x) > f_j(x^*) \text{ para algún } i \in [1, k])$$

De acuerdo con Deb [13], si se calcula la dominancia de un conjunto finito de soluciones se espera obtener un subconjunto de soluciones con pares de soluciones no dominadas entre sí. Dicho conjunto tendrá la propiedad de dominar al resto de soluciones del espacio que pertenezcan a él y recibirá el nombre de “conjunto de soluciones no dominadas”. Cuando el conjunto finito de soluciones es todo el espacio de soluciones, el conjunto de soluciones no dominadas que se obtenga de él se conoce como conjunto óptimo de Pareto. Por otro lado, al conjunto que contiene todas las funciones objetivo de las soluciones del conjunto óptimo de Pareto se le conoce como frente de Pareto. Las soluciones no dominadas que se obtengan durante la resolución de un problema multiobjetivo estarán a la disposición del tomador de decisiones para seleccionar aquella que mejor se adapte a las preferencias de la organización.

## 2.4 Problema de Selección de Cartera de Proyectos

El problema de selección de cartera de proyectos es un problema de optimización multiobjetivo que suele presentarse en el entorno organización y cuya formulación estpa basada en la del problema de selección de portafolio (*Portfolio Selection Problem*, PSP por sus siglas en inglés), propuesto en 1952 [16]. El problema se hace presente cuando, con el propósito de conseguir oportunidades de crecimiento o tratar de garantizar su estabilidad, las empresas invierten en un conjunto de proyectos cuyo costo total excede los recursos de los que la organización dispone y deberá entonces seleccionarse aquella combinación de proyectos que le garantice el máximo beneficio empleando los recursos disponibles.

En el contexto del problema, un proyecto representa un proceso temporal único e irrepetible que tiene un conjunto específico de objetivos. Una cartera estará conformada por un conjunto de proyectos que comparten recursos durante un determinado periodo de tiempo [17].

El modelo matemático del problema de selección de cartera de proyectos consta de los siguientes elementos: sean  $n$  el número de proyectos a considerar,  $C$  el presupuesto total disponible,  $O$  el número de objetivos,  $c_i$  el costo de inversión para el proyecto  $i$ ,  $b_{ij}$  el beneficio que representa la inversión en el proyecto  $i$  al objetivo  $j$ ,  $K$  y  $M$  el número de áreas administrativas y regiones geográficas a considerar respectivamente,  $A_k^{min}$  y  $A_k^{max}$  los límites inferior y superior del presupuesto asignado al área  $k$ ,  $R_m^{min}$  y  $R_m^{max}$  los límites inferior y superior del presupuesto asignado a la región  $m$ ,  $a_i$  y  $b_i$  valores que representan el área y la región a las que pertenece el proyecto  $i$ .  $x = (x_1, x_2, \dots, x_n)$  es un vector binario que especifica los proyectos incluidos en una determinada cartera, de manera que si  $x_i = 1$ , el proyecto  $i$  se encuentra dentro de ella y no lo hará en caso contrario. Con todos los elementos antes mencionados, puede realizarse la siguiente formulación del problema:

$$\text{Maximizar } Z = (z_1, z_2, \dots, z_O) \quad (1)$$

Donde:

$$z_j = \sum_{i=1}^n b_{ij} x_i; \quad j = 1, 2, \dots, O \quad (2)$$

Sujeto a:

$$\sum_{i=1}^n c_i x_i \leq C \quad (3)$$

$$A_k^{min} \leq \sum_{i=1, a_i=k}^n c_i x_i \leq A_k^{max}; \quad k = 1, 2, \dots, K \quad (4)$$

$$R_m^{min} \leq \sum_{i=1, b_i=m}^n c_i x_i \leq R_m^{max}; \quad m = 1, 2, \dots, M \quad (5)$$

$$x_i \in \{0, 1\} \text{ para } i = 1, 2, \dots, n \quad (6)$$

La función objetivo maximiza las contribuciones a cada uno de los objetivos (1). El valor de los objetivos se calcula mediante la adición de la contribución que cada uno de los proyectos presentes en la cartera representa para cada objetivo (2). La restricción de presupuesto (3) asegura que la suma de los costos de inversión de los proyectos seleccionados no exceda el presupuesto total. De manera similar, las restricciones de área (4) y región (5)

aseguran que la suma de los costos de los proyectos pertenecientes a determinada área y región se encuentren dentro de los límites de presupuesto establecidos. Una cartera estará representada mediante un vector de variables de decisión que indicará los proyectos presentes en ella (6).

## 2.5 El modelado de la incertidumbre

La vida de todo ser humano se encuentra plagada de procesos de toma de decisiones. Desde el orden de las acciones a realizar durante su rutina matutina hasta la elección de una hora para dormir, el ser humano lleva a cabo su día navegando entre un sinfín de alternativas.

Habitualmente, la elección de alternativas en procesos tan simples como el orden en que se colocan las prendas a la hora de vestirse suelen realizarse de manera inconsciente y casi automática. Sin embargo, existen situaciones en las que las alternativas involucran factores complejos que dificultan el proceso de toma de decisiones (*e.g.*: la contradicción entre objetivos, ambigüedades en los requerimientos o la mala especificación de los mismos), siendo uno de los más comunes la presencia de la incertidumbre.

La incertidumbre suele presentarse a través de sus distintas facetas durante la formulación de una gran variedad de problemas de optimización debido a su inherente presencia en el diseño, operación y control de sistemas complejos que representan o se inspiran en situaciones del mundo real [1] [4]. Su presencia puede estar directamente relacionada con una variedad de factores, como [18] [19]:

- El inconveniente de no contar con cantidades suficientes de información confiable.
- La falta de predicciones confiables de características, propiedades y comportamiento de sistemas complejos ante acciones internas o externas.
- Restricciones y objetivos pobremente definidos.
- La dificultad de formalizar ciertos factores y criterios y la necesidad de tomar en cuenta información cualitativa.

Tomar en consideración el factor de la incertidumbre durante el diseño y construcción del modelo matemático de un problema de optimización permite adaptarlo a condiciones más apegadas a los problemas del mundo real [1]; incrementando así el grado de credibilidad de los resultados obtenidos durante su resolución y facilitando el análisis de los mismos para posibles procesos de toma de decisiones. La incertidumbre puede presentarse en distintos formatos, algunos de ellos son [20]:

- Probabilidades.
- Peso de las decisiones (preferencias).
- Medidas de riesgo explícitas.
- Números difusos.
- Escenarios.

A lo largo de la historia, uno de los enfoques más ampliamente utilizados para modelar la incertidumbre es aquel basado en la teoría de la probabilidad [8], siendo de gran utilidad para modelar aquellas situaciones en las que la fuente de la incertidumbre es el azar [5]. Sin embargo, la incertidumbre puede provenir de fuentes diversas con distintas características para las cuales el enfoque probabilístico podría no ser el más adecuado [1]. En muchas ocasiones, durante el proceso de análisis y modelado de problemas del mundo real, el hecho de conocer si determinada condición se presenta o no, carece de utilidad práctica y es necesario conocer *qué tanto* de dicha condición está presente.

Algunas características o condiciones que se presentan en situaciones del mundo real pueden resultar difíciles de describir mediante cuantificaciones numéricas clásicas. Por ello, el ser humano suele describirlas de manera difusa de acuerdo a su percepción, empleando expresiones en lenguaje natural (conocidas como estados lingüísticos) [8]. Descripciones tan básicas como “grande”, “rápido”, o “muy frío” son ejemplos claros de estados lingüísticos empleados cotidianamente para ello. No obstante, dichos estados son una fuente clara de incertidumbre, ¿Qué tamaño tiene algo considerado grande? ¿A qué velocidad se desplaza algo rápido? ¿A qué temperatura se encuentra algo considerado muy frío? Son algunas preguntas que surgen dada la ambigüedad de las descripciones.

Actualmente, existen mecanismos que permiten modelar eventos ambiguos de manera natural y apegada al razonamiento humano haciendo uso de la teoría de conjuntos difusos y la lógica difusa [5]. Ambos basan su funcionamiento en la generación de reglas que permiten caracterizar a los estados lingüísticos, estableciendo umbrales para cada uno de ellos. Permitiendo así la incorporación del factor de incertidumbre al modelado de problemas de optimización complejos.

El presente trabajo de tesis se centra en modelar y establecer una estrategia de resolución de la variante multiobjetivo del problema de selección de cartera de proyectos tomando en consideración la incertidumbre que puede presentarse tanto en las restricciones como en las funciones objetivo. El tipo de incertidumbre que se aborda en el mismo, se expresa mediante proposiciones con el estado lingüístico “cerca a” o “cerca de” [5] usado para denotar costos de inversión y posibles beneficios o retorno.

## **2.6 Elementos de la teoría de números difusos**

En esta sección se describen algunos de los conceptos fundamentales de la teoría de números difusos que permiten la incorporación del factor de incertidumbre a un modelo matemático para la variante del problema de selección de cartera de proyectos que se aborda en este trabajo de tesis. Así mismo, se describen los elementos básicos que permiten su incorporación y resolución mediante un algoritmo genético basado en el principio de no dominancia.

### 2.6.1 Conjuntos difusos

En la teoría de conjuntos, un conjunto “clásico” se define como una colección (finita o infinita) de elementos (también llamados miembros) distintos entre sí que comparten ciertas características [8]. Existen tres formas básicas de representación de conjuntos [5]:

1. Listando sus elementos: suele emplearse para conjuntos finitos. Un conjunto  $A$  cuyos miembros son  $a_1, a_2, a_3, \dots, a_n$  suele denotarse como:

$$A = \{a_1, a_2, a_3, \dots, a_n\} \quad (7)$$

2. Por una regla: un conjunto  $A$  está conformado por los elementos que cumplen determinada condición. Por ejemplo:

$$A = \{x|P(x)\} \quad (8)$$

Donde  $X$  es el conjunto universal y  $x \in X$ .

Bajo esta representación (ecuación 8) se denota que el conjunto  $A$  está conformado por todos aquellos elementos cuya evaluación en la proposición  $P(x)$  es verdadera.

3. Por una función: un conjunto se representa mediante una “función característica” que declara qué elementos de  $X$  forman parte de un conjunto y cuáles no. Por ejemplo, un conjunto  $A$  puede definirse mediante su función característica  $X_A$  de la siguiente manera:

$$X_A = \begin{cases} 1 & \text{para } x \in A \\ 0 & \text{para } x \notin A \end{cases} \quad (9)$$

Donde la función característica mapea los elementos de  $X$  al conjunto  $[1,0]$ , de forma se considera que aquellos elementos en que  $X_A(x) = 1$  serán miembros de  $A$  y no lo serán en caso contrario.

De forma similar, en el caso de los conjuntos difusos, la función característica permite establecer varios grados de pertenencia a sus elementos.

Si  $X$  es una colección de objetos denotados genéricamente como  $x$ , entonces un conjunto difuso  $\tilde{A}$  en  $X$  es un conjunto de pares ordenados [21] [22]:

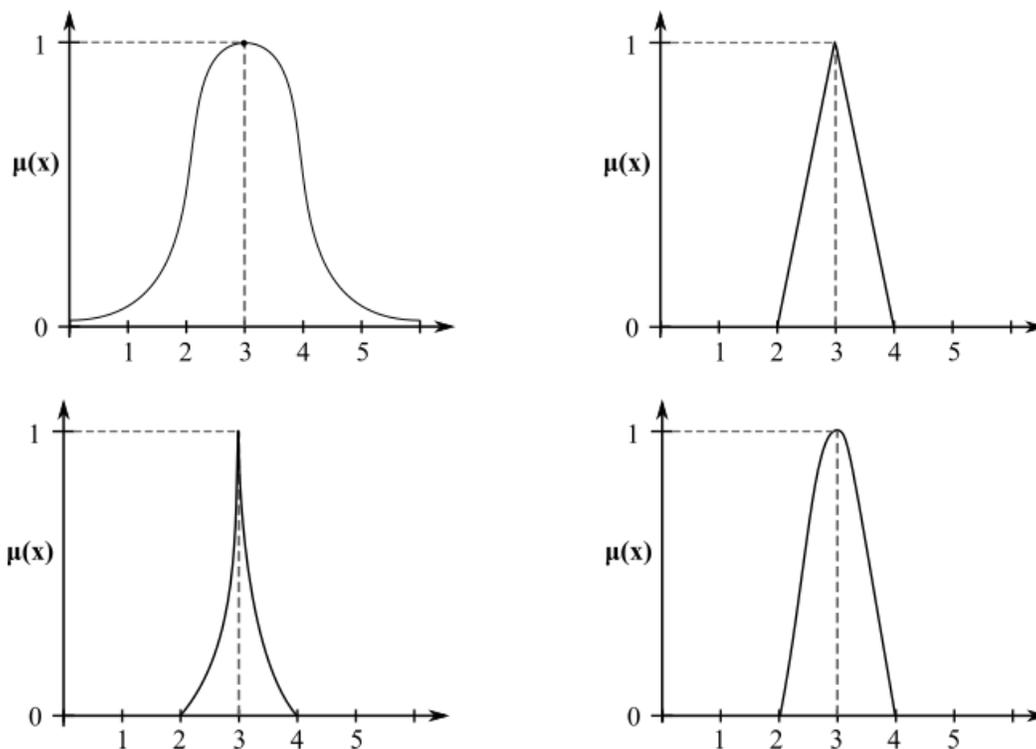
$$\tilde{A} = \{x, \mu_{\tilde{A}}(x)|x \in X\} \quad (10)$$

A la función  $\mu_{\tilde{A}}(x)$  se le conoce como función (o grado) de pertenencia (o membresía) de  $x$  en  $\tilde{A}$  que mapea a  $X$  en un espacio de membresía  $M$  [22]. El rango más comúnmente usado para las funciones de pertenencia de los conjuntos difusos es el intervalo unitario  $[0,1]$ .

Los conjuntos difusos permiten modelar la incertidumbre en situaciones en las que los límites empleados para clasificar un determinado evento u objeto (*i.e.*: “pesado”, “joven”, “alto”) no se encuentran claramente definidas asignando un significado matemático a expresiones del lenguaje natural, facilitando su manipulación y combinación con otros conceptos matemáticos [23].

La representación de distintas situaciones mediante conjuntos difusos no depende únicamente del concepto, sino también dependerá del contexto en que presenten dichas situaciones [5], por ejemplo, el concepto de “alta velocidad” tiene distintas percepciones en los contextos de ciclismo y automovilismo. El contexto tendrá influencia directa en la función de membresía a seleccionar, generando la posibilidad de generar distintas representaciones del mismo evento mediante la utilización de distintas funciones de membresía.

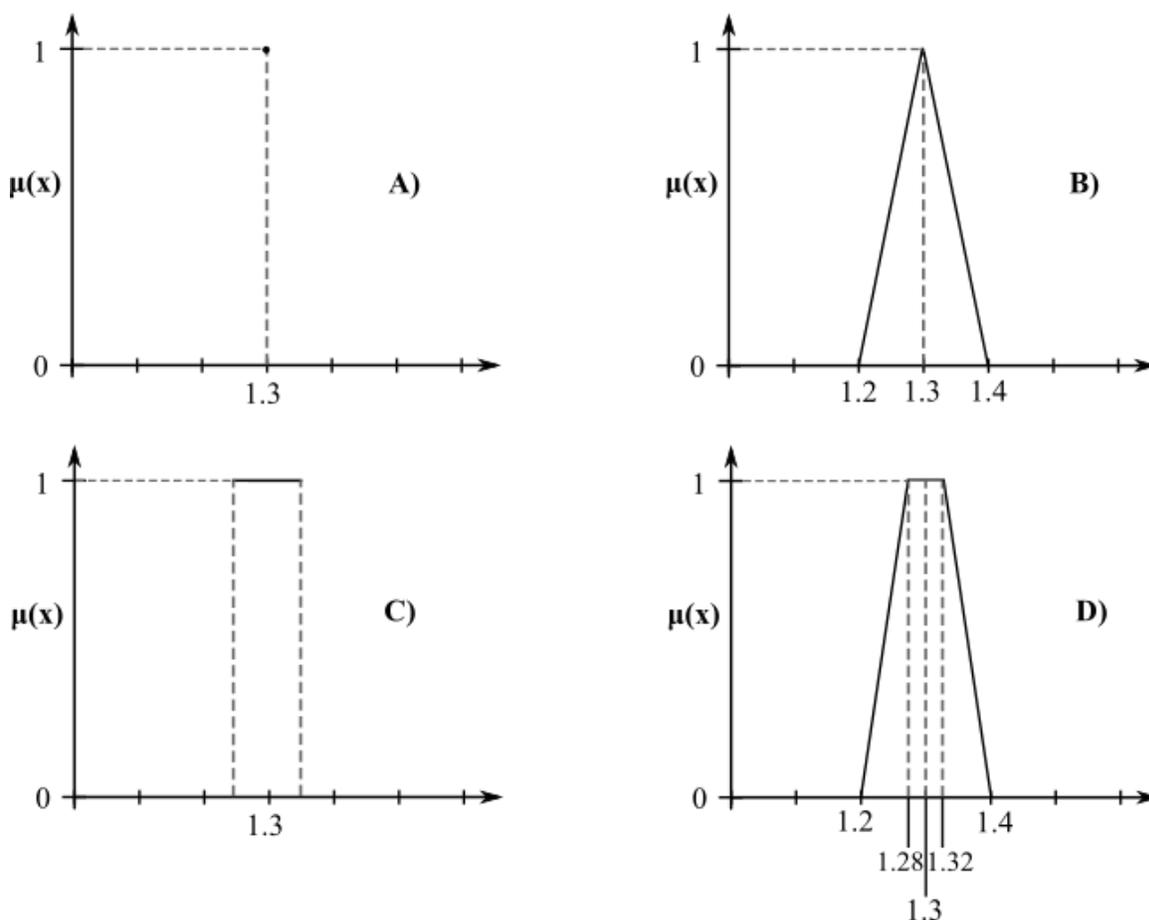
Un uso común de los conjuntos difusos es la representación de cantidades cuyo valor no está claramente definido. A dicha aplicación se le conoce como “números difusos” y son ampliamente usados en una gran variedad de contextos [5], ya que permiten representar nuestra concepción habitual de los números aproximados o cantidades imprecisas [1]. En el contexto que se aborda en este trabajo de tesis, por ejemplo, se utiliza para modelar la incertidumbre presente en los costos de inversión y en el posible retorno que dicha inversión puede generar; incertidumbre que puede provenir de factores como la volatilidad de los mercados, o las tasas de cambio.



**Figura 1.** Ejemplos de funciones de membresía que pueden usarse en distintos contextos para caracterizar conjuntos difusos de números reales cercanos a 3.

En la Figura 1 se muestran cuatro posibles representaciones de la misma situación. En ellas, es posible apreciar una notoria diferencia en la forma de cada una de las gráficas. Dicha diferencia se debe al uso de distintas funciones de membresía. La factibilidad usar una forma en particular sólo podrá determinarse con base en el contexto de su aplicación. Sin embargo, el usar una u otra forma no tiene un efecto particularmente fuerte al modelar la gran mayoría de las situaciones [5], por lo que suelen emplearse formas sencillas como la triangular o trapezoidal.

Un número difuso  $A$  representa una generalización de un número real  $r$  y proporciona un mecanismo que permite representar valores aproximados expresados con proposiciones como “números cercanos a” o “números ubicados en el intervalo de”. En donde únicamente  $A(r) = 1$  [23]. Como se mencionó anteriormente, las funciones de membresía que se seleccionen deberán estar directamente influenciadas por el tipo de incertidumbre que se desea modelar.



**Figura 2.** Representación gráfica de distintos tipos de valores.

En la Figura 2 pueden observarse distintos ejemplos de representación de valores reales y difusos. En dicha figura:

- A) Es un valor entero: 1.3.
- B) Es un número difuso expresando la proposición “cerca de 1.3”.
- C) Es un intervalo cerrado de números reales [1.25, 1.35].
- D) Es un número difuso con una sección plana (número trapezoidal).

En las siguientes secciones se describen las características generales de las funciones de membresía. Además, se describen dos de las estrategias más utilizadas para representar números difusos.

### 2.6.2 Números difusos generalizados

Un número difuso generalizado, denotado comúnmente como  $\tilde{A}$ , es un subconjunto cualquiera del conjunto de números reales  $\mathbb{R}$  cuya función de membresía  $\mu_{\tilde{A}}(x)$  satisface las condiciones siguientes [21]:

1.  $\mu_{\tilde{A}}(x)$  es un mapeo continuo de  $\mathbb{R}$  en el intervalo [0,1].
2.  $\mu_{\tilde{A}}(x) = 0, -\infty < x < a$ .
3.  $\mu_{\tilde{A}}(x) = L(x)$ , es estrictamente creciente para  $[a, b]$ .
4.  $\mu_{\tilde{A}}(x) = w$ , para  $b < x < \alpha$ .
5.  $\mu_{\tilde{A}}(x) = R(x)$ , es estrictamente decreciente para  $[\alpha, \beta]$ .
6.  $\mu_{\tilde{A}}(x) = 0$ , para  $\beta < x < \infty$ .

Donde  $0 < w < 1, a, b, \alpha, \beta$  son números reales.

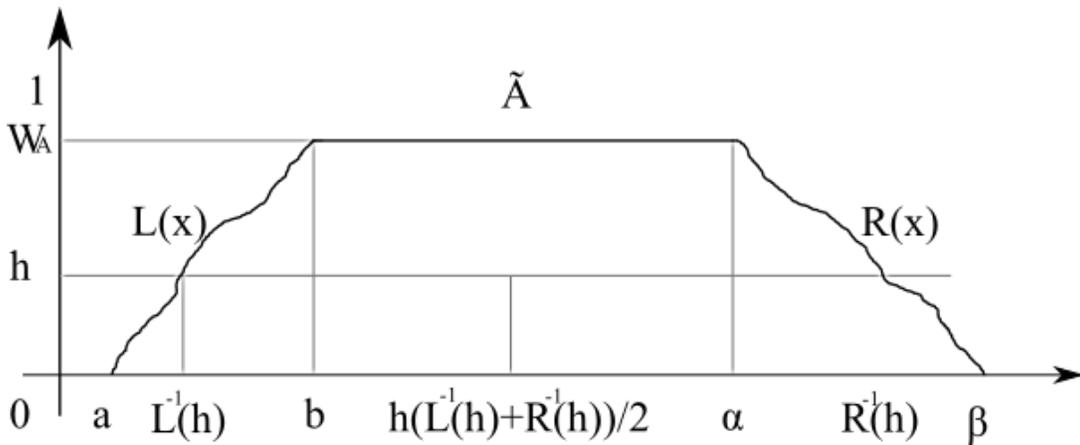


Figura 3. Ejemplo de número difuso generalizado.

En la Figura 3 es posible observar los elementos que componen un número difuso generalizado. De forma general, es posible denotar un número difuso generalizado "A" como:  $\tilde{A} = (a, b, \alpha, \beta, w)_{LR}$ . Cuando  $w=1$ , el número se denota como  $\tilde{A} = (a, b, \alpha, \beta)_{LR}$ .

Es posible clasificar los números difusos de acuerdo a las características de los elementos que los conforman. Por lo regular, el tipo de un número difuso generalizado está directamente determinado por la forma de su función de pertenencia. En el presente trabajo, a pesar de centrarse en números difusos del tipo triangular, es preciso mencionar los números difusos trapezoidales debido a sus características y propiedades comunes. A continuación, se describen las características de ambos tipos de números difusos.

### 2.6.3 Números difusos trapezoidales

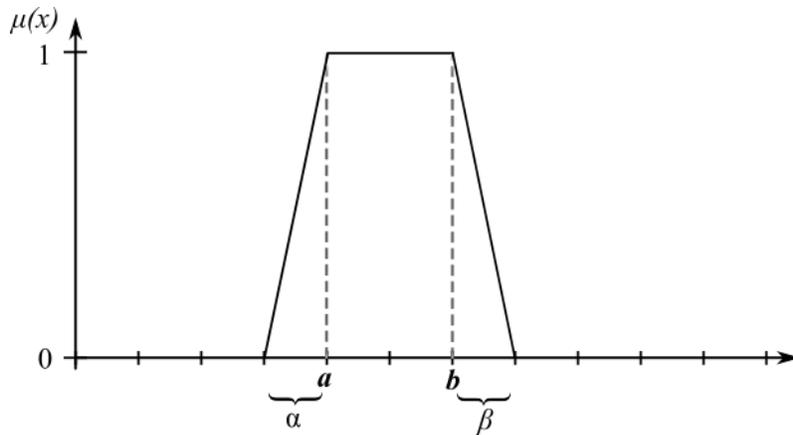
El uso de números difusos trapezoidales suele ser apropiado para modelar situaciones en las que un intervalo de valores no puede ser definido con precisión [23]. Para modelar dicho intervalo como un *intervalo difuso* es necesario establecer un intervalo [a, b] cuyos miembros tomarán el valor de 1.

Los números difusos trapezoidales se caracterizan porque su función de membresía describe la forma de un trapecio, ya que los valores de las funciones izquierda y derecha ( $L(x)$  y  $R(x)$ , respectivamente) que lo caracterizan son líneas rectas. Un número difuso trapezoidal  $A$  se denota como [21]:

$$\tilde{A} = (a, b, \alpha, \beta) \quad (11)$$

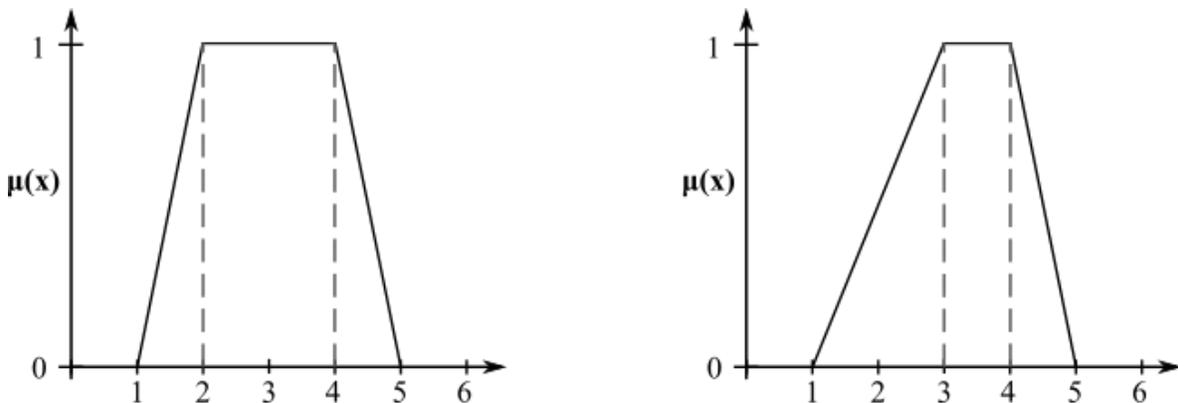
Y su función de membresía se define como:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x \leq a_{\min} \\ \frac{x - a_{\min}}{\beta_1 - a_{\min}}, & x \in (a_{\min}, \beta_1) \\ \frac{a_{\max} - x}{a_{\max} - \beta_2}, & x \in (\beta_2, a_{\max}) \\ 0, & x \geq a_{\max} \end{cases} \quad (12)$$



**Figura 4.** Representación gráfica de un número difuso trapezoidal.

En la Figura 4 es posible observar los elementos que componen a un número difuso trapezoidal: la sección plana  $[a, b]$  conocida como “intervalo difuso” [5] y los componentes  $\alpha$  y  $\beta$ .



**Figura 5.** Ejemplos de representación de intervalos mediante números trapezoidales.

Como se mencionó anteriormente, la forma que adopte una representación de número difuso está directamente relacionada con el contexto de la situación en la que se produzca la incertidumbre. En la Figura 5, se observan dos representaciones de intervalos difusos: la gráfica del lado izquierdo representa al intervalo  $[2, 4]$ , mientras que la del lado derecho representa al intervalo  $[3, 4]$ . Como es posible observar, la función de pertenencia trapezoidal puede no siempre tener forma simétrica.

Este tipo de números suele utilizarse para modelar la incertidumbre expresada en proposiciones del tipo “un valor ubicado aproximadamente entre  $\_y\_$ ” [23]. En este sentido, es posible expresar los números difusos que se observan en las gráficas de la Figura 5 con proposiciones como “valores ubicados aproximadamente entre 2 y 4” y “valores que se encuentran aproximadamente entre 3 y 4” respectivamente.

### 2.6.4 Números difusos triangulares

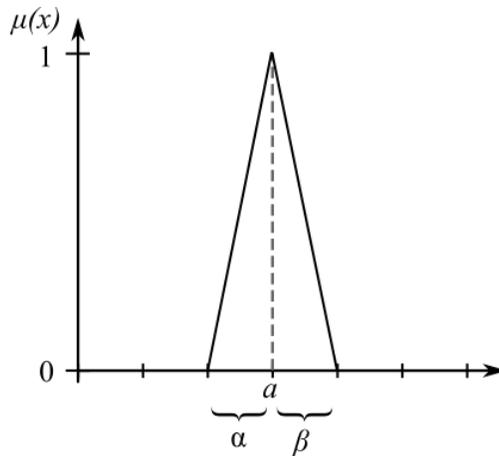
Los números difusos de tipo triangular son considerados el mecanismo más simple usado para modelar la incertidumbre ya que se conforma únicamente de tres elementos [5]. Además, puede considerarse a este tipo de números como un caso especial de números difusos trapezoidales en los que los valores de  $a$  y  $b$  son iguales [21].

Suele recurrirse a los números difusos triangulares para modelar la incertidumbre expresada mediante proposiciones del tipo “cercano a” o “cerca de” [5]. Un número difuso triangular  $A$  se denota como [21]:

$$\tilde{A} = (a, \alpha, \beta) \quad (13)$$

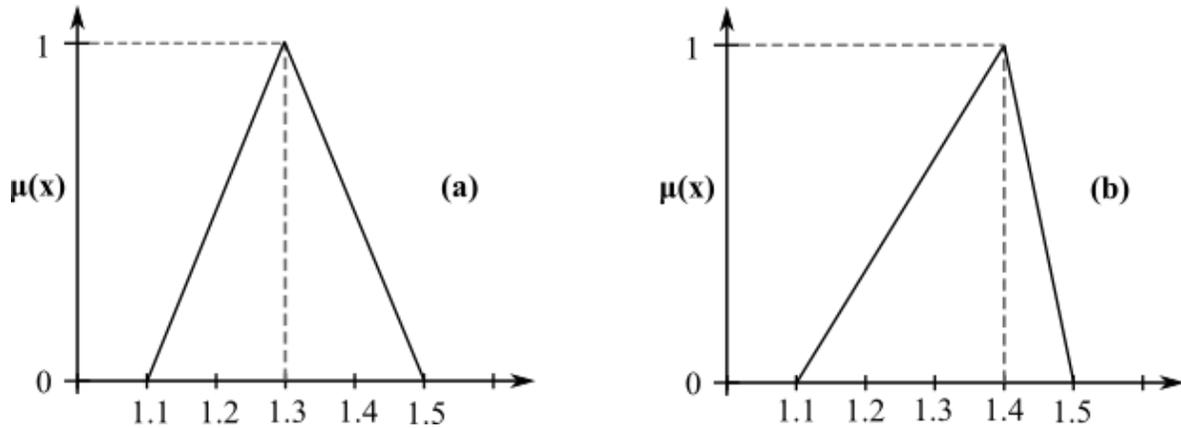
Y su función de membresía se define como:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x \leq a_{\min} \\ \frac{x - a_{\min}}{\beta - a_{\min}}, & x \in (a_{\min}, \beta) \\ \frac{a_{\min} - x}{a_{\min} - \beta}, & x \in (\beta, a_{\min}) \\ 0, & x \geq a_{\max} \end{cases} \quad (14)$$



**Figura 6.** Representación gráfica de un número difuso triangular.

En la Figura 6 es posible observar los elementos que componen a un número difuso triangular:  $a$ , que denota un valor modal de un determinado conjunto difuso [18] y los componentes  $\alpha$  y  $\beta$ .



**Figura 7.** Representación gráfica de un número difuso triangular.

En la Figura 7 se muestran las representaciones gráficas de dos números difusos triangulares. Al comparar ambas gráficas es posible observar una notoria falta de simetría en la gráfica (b). Dicha falta de simetría puede ser directamente causada por la contextualización de los eventos o situaciones modeladas.

Los números difusos triangulares que pueden observarse en la Figura 7 pueden ser descritos usando expresiones en lenguaje natural tales como “cerca de 1.3” o “cercano a 1.3” para el inciso (a) y “aproximadamente 1.4” o “alrededor de 1.4” para el inciso (b) (siendo cada una de las proposiciones anteriores válidas en cualquiera de los dos casos).

En este trabajo, se emplean los números difusos triangulares para modelar la incertidumbre de los parámetros (presupuestos, costos de inversión y beneficio/retorno) de la variante multiobjetivo del problema de selección de cartera de proyectos. En las siguientes secciones se detallan los conceptos matemáticos y relacionales en los que se sustentan las estrategias de modelado, adaptación y resolución propuestas en el mismo.

## 2.7 Aritmética de números difusos

Al igual que en la teoría de la aritmética clásica, en la que se establece un conjunto de reglas que definen a las operaciones realizables sobre los números, en la teoría de números difusos, existen reglas (aún no estandarizadas) que definen operaciones aritméticas en el espacio de los números difusos. Dichas operaciones pueden resultar especialmente complicadas si los números difusos involucrados fueron modelados usando funciones de pertenencia complejas. Por ello es recomendable, siempre que sea posible, modelar la incertidumbre empleando

funciones de pertenencia trapezoidales o triangulares, ya que su número reducido de componentes facilita la aplicación de reglas que permitan realizar dichas operaciones.

Dado que este trabajo se enfoca en incertidumbre modelada mediante números difusos triangulares y que estos son un caso especial de números trapezoidales [21], se hace uso de una adaptación de las reglas que permiten realizar operaciones aritméticas entre números difusos trapezoidales.

En las siguientes dos secciones se definen las operaciones aritméticas entre números trapezoidales utilizadas en este trabajo y su adaptación a los números triangulares.

### 2.7.1 Operaciones aritméticas entre números trapezoidales

Si  $\tilde{A} = (a_1, b_1, \alpha_1, \beta_1)$  y  $\tilde{B} = (a_2, b_2, \alpha_2, \beta_2)$  son dos números difusos trapezoidales. Es posible definir operaciones aritméticas entre mediante las siguientes reglas [21]:

- Suma:

$$\tilde{A} + \tilde{B} = (a_1 + a_2, b_1 + b_2, \alpha_1 + \alpha_2, \beta_1 + \beta_2).$$

- Resta:

$$\tilde{A} - \tilde{B} = (a_1 - b_2, b_1 - a_2, \alpha_1 + \beta_2, \beta_1 + \alpha_2).$$

- Multiplicación:

$$\tilde{A} * \tilde{B} = (a_1 * a_2, b_1 * b_2, |a_1| * \alpha_2 + \alpha_1 * |a_2|, |b_1| * \beta_2 + \beta_1 * |b_2|).$$

- División:

$$\frac{\tilde{A}}{\tilde{B}} = \left( \frac{a_1}{b_2}, \frac{b_1}{a_2}, \frac{|a_1| * \beta_2 + \alpha_1 * |b_2|}{b_2^2}, \frac{|a_1| * \alpha_2 + \beta_1 * |a_2|}{a_2^2} \right).$$

### 2.7.2 Operaciones aritméticas entre números triangulares

Como se mencionó anteriormente, los números difusos triangulares representan casos especiales de números difusos trapezoidales en los que los límites del intervalo difuso tienen el mismo valor, es decir:  $a=b$ . Con base en dicha propiedad, es posible reformular las reglas que definen las operaciones entre números trapezoidales descritas en la sección anterior para adaptarlas a las características de los números difusos triangulares.

Si se considera que  $\tilde{A} = (a_1, \alpha_1, \beta_1)$  y  $\tilde{B} = (a_2, \alpha_2, \beta_2)$  son dos números difusos de tipo triangular. Las operaciones aritméticas entre ellos se definen de la siguiente manera:

- Suma:  

$$\tilde{A} + \tilde{B} = (a_1 + a_2, \alpha_1 + \alpha_2, \beta_1 + \beta_2).$$
- Resta:  

$$\tilde{A} - \tilde{B} = (a_1 - a_2, \alpha_1 + \beta_2, \beta_1 + \alpha_2).$$
- Producto:  

$$\tilde{A} * \tilde{B} = (a_1 * a_2, |a_1| * \alpha_2 + \alpha_1 * |a_2|, |a_1| * \beta_2 + \beta_1 * |a_2|).$$
- División:  

$$\frac{\tilde{A}}{\tilde{B}} = \left( \frac{a_1}{a_2}, \frac{|a_1| * \beta_2 + \alpha_1 * |a_2|}{a_2^2}, \frac{|a_1| * \alpha_2 + \beta_1 * |a_2|}{a_2^2} \right).$$

## 2.8 Comparación de números difusos

Además del conjunto de reglas que define cómo se realizan las operaciones aritméticas básicas sobre los números difusos triangulares, existen diversas estrategias que permiten realizar otras de las operaciones tradicionalmente aplicadas a números reales sobre números difusos. Una de las operaciones más elementales es la comparación, lo cual permite establecer una relación de orden entre los números difusos.

Desde la concepción de los números difusos se han propuesto diversas estrategias para su comparación. La inmensa mayoría de dichas estrategias consisten en aplicar algún proceso matemático (similar a un proceso de normalización) a los valores que describen al número difuso para transformarlo en un valor real (comúnmente conocido como índice) que pueda ser comparado con otro valor de igual naturaleza. Al proceso usado para obtener un índice que describa la magnitud de un número difuso se le suele denominar “*defuzzification*” o “*defuzzificación*” [5].

En esta sección, se describe el proceso de *defuzzificación* usado en este trabajo.

### 2.8.1 Graded Mean Integration

*Graded Mean Integration* (comúnmente abreviada como GMI) o Integración de Media Graduada [24] es un método de normalización de números difusos (usualmente llamado método de *defuzzificación*) empleado para facilitar su comparación y establecer una relación de orden entre ellos. El número difuso mayor será aquel que cuente con más alto valor de GMI.

El valor de GMI de un número difuso  $\tilde{A}$ , denotado como  $P(\tilde{A})$  se obtiene mediante la siguiente fórmula:

$$P(\tilde{A}) = \frac{\int_0^w h \left( \frac{L^{-1}(h) + R^{-1}(h)}{2} \right) dh}{\int_0^w h dh} \quad (15)$$

En el caso particular de los números difusos trapezoidales, la ecuación (15) puede simplificarse de la siguiente manera:

$$P(\tilde{A}) = \frac{3a + 3b + \beta - \alpha}{6} \quad (16)$$

De igual manera, dado que los números difusos triangulares son un caso especial de números trapezoidales en que  $a=b$ , la ecuación (16) puede simplificarse como:

$$P(\tilde{A}) = \frac{6a + \beta - \alpha}{6} \quad (17)$$

## 2.8.2 Relación de orden entre números difusos

Con base en los conceptos antes descritos, es posible establecer la relación de orden entre dos números difusos triangulares. En este trabajo, se hace uso del mecanismo de *defuzzificación* GMI para transformar los valores difusos y poder así comparar sus magnitudes. Si  $\tilde{A}_1$  y  $\tilde{A}_2$  son dos números difusos triangulares, la relación de orden entre ellos aplicando el

- $\tilde{A}_1 > \tilde{A}_2$  si y solo si  $P(\tilde{A}_1) > P(\tilde{A}_2)$ .
- $\tilde{A}_1 < \tilde{A}_2$  si y solo si  $P(\tilde{A}_1) < P(\tilde{A}_2)$ .
- $\tilde{A}_1 = \tilde{A}_2$  si y solo si  $P(\tilde{A}_1) = P(\tilde{A}_2)$ .

## 2.8.3 Dominancia entre vectores difusos

Con base en la definición de dominancia [13], es posible establecer una relación de dominancia entre dos vectores difusos:  $X = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$  y  $Y = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n\}$ , donde  $\tilde{x}_i$  y  $\tilde{y}_i$  son números difusos triangulares, empleando el mecanismo de comparación GMI. Se dice que:  $X > Y$  si y solo si  $P(\tilde{x}_i) \geq P(\tilde{y}_i) \forall i \in \{1, 2, \dots, n\}$  y  $P(\tilde{x}_i) > P(\tilde{y}_i)$  para al menos un valor de  $i \in \{1, 2, \dots, n\}$  [25].

## 2.9 Metaheurísticas poblacionales

Las metaheurísticas poblacionales son estrategias comúnmente empleadas en la resolución de problemas de optimización que basan su funcionamiento, como su nombre lo indica, en la optimización de un conjunto de soluciones (conocido como población), a diferencia de otras metaheurísticas que buscan optimizar una única solución (metaheurísticas trayectoriales). Existen diversos tipos de heurísticas poblacionales, las más comunes son aquellas inspiradas en los procesos de la naturaleza, como la evolución (algoritmos evolutivos, algoritmos genéticos) o el comportamiento de grupos de animales (algoritmos de optimización de colonia de hormigas, algoritmos de optimización de enjambre) [26]. En ellos, al conjunto de individuos que conforma la población se le asigna una función de aptitud y de acuerdo al avance del proceso de búsqueda de soluciones, sólo se conservarán aquellos individuos con mejor aptitud, buscando así alcanzar soluciones óptimas.

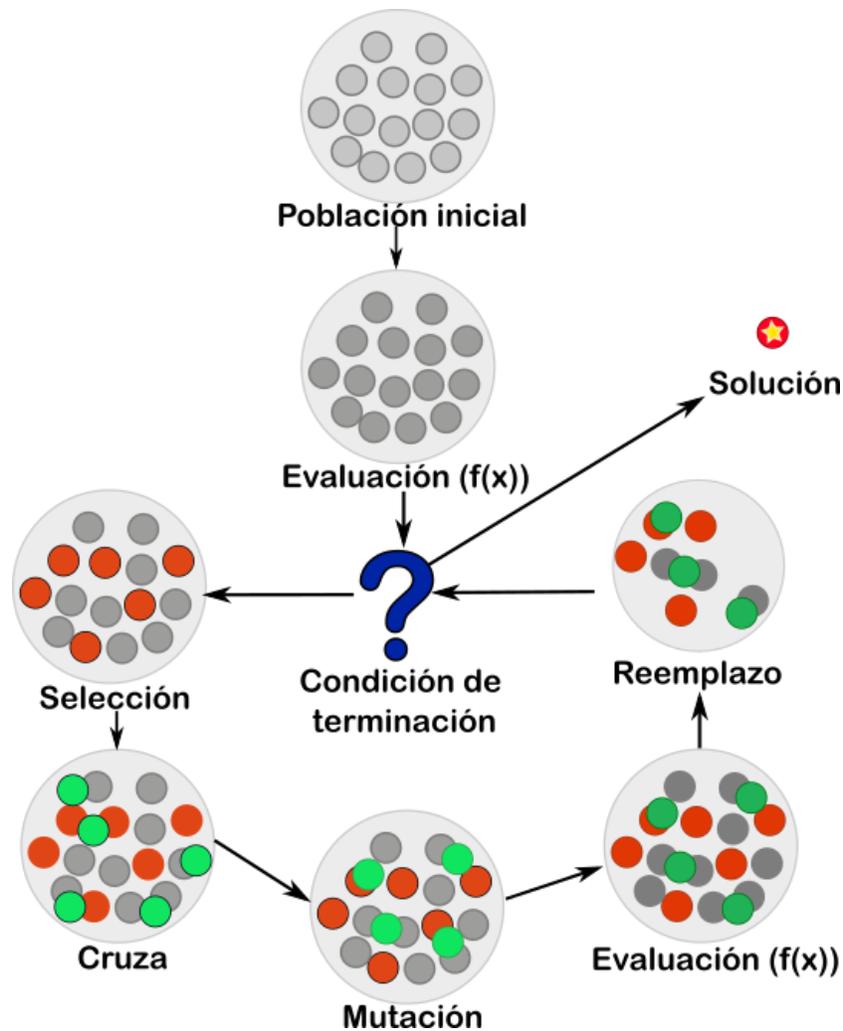
### 2.9.1 Algoritmos genéticos

Los algoritmos genéticos (*Genetic Algorithms*, G.A. por sus siglas en inglés) constituyen un grupo de algoritmos evolutivos de optimización que surgió en la década de los 60 de la mano de John Holland [27]. Este tipo de algoritmos recibe su nombre porque su concepción estuvo inspirada en los procesos de evolución y selección natural, lo cual establece un método de optimización general que es aplicable a un amplio conjunto de problemas.

Los algoritmos genéticos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (mutaciones y recombinaciones genéticas), así como también a una selección de acuerdo con algún criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados.

Un algoritmo genético típico consta de los siguientes elementos [28]:

1. Un número de aproximaciones a la solución del problema (población).
2. Una manera de calcular que tan buena o mala es cada una de las soluciones que conforman la población. (evaluación y selección).
3. Un método para mezclar fragmentos de las mejores soluciones para formar nuevas (cruza).
4. Un operador de mutación para evitar pérdidas permanentes de diversidad entre las soluciones (mutación).



**Figura 8.** Diagrama general del funcionamiento de un algoritmo genético.

## 2.10 Pruebas no paramétricas

En la resolución de problemas mediante algoritmos computacionales, al igual que en diversos campos de la ciencia, es necesario contar con mecanismos que permitan comparar los resultados obtenidos a través de un proceso de experimentación, para ello, en este trabajo se emplean pruebas no paramétricas.

Las pruebas no paramétricas son un tipo de pruebas de hipótesis cuya aplicación no requiere que la distribución de la población a tratar o sus parámetros tengan una forma determinada, por lo que se les denominan como pruebas libres de distribución (*distribution free*) [29]. En la mayor parte de ellas los resultados estadísticos se derivan únicamente de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión y son particularmente útiles en aquellas situaciones en las que se trabaja con muestras

pequeñas ( $n < 10$ ) en las que se desconoce si es válido suponer la normalidad de los datos [30].

### 2.10.1 Prueba de rangos con signo de Wilcoxon

La prueba de los rangos con signo de Wilcoxon es una prueba no paramétrica para comparar el rango medio de dos muestras relacionadas y determinar si existen diferencias entre ellas [31]. Suele utilizarse como alternativa a la prueba t de Student cuando no se puede suponer la normalidad de dichas muestras [30]. En este tipo de prueba se dispone de  $n$  pares de observaciones, denotadas como  $(x_i, y_i)$  y su objetivo radica en comprobar si puede dictaminarse que los valores de  $x_i$  e  $y_i$  son o no iguales.

Suponiendo que  $z_i = y_i - x_i$ , entonces los valores de  $z_i$  son independientes y que los valores de  $x_i$  e  $y_i$  tienen una misma distribución continua y simétrica respecto a  $\theta$ . El procedimiento a seguir para efectuar la prueba se describe a continuación [29] [30] [31]:

Se considera que la hipótesis nula:  $H_0: \theta = 0$ , indicando que los valores de  $x_i$  e  $y_i$  originales son en cierto sentido del mismo tamaño. Para verificar dicha hipótesis, se realiza primero el ordenamiento de los valores absolutos de  $z_i$  (i.e.:  $|z_1|, |z_2|, \dots, |z_n|$ ) y de acuerdo a ello, se les asigna un rango  $R_i$ .

Posteriormente, se procede a calcular el estadístico de la prueba de signos de Wilcoxon ( $W^+$ ), de la siguiente forma:

$$W^+ = \sum_{z_i > 0} R_i \quad (18)$$

La distribución del valor de  $W^+$  puede consultarse en tablas para determinar si la hipótesis nula se acepta o no.

Por otro lado, en algunas ocasiones, suele emplearse esta prueba para comparar las diferencias entre dos muestras de datos tomados antes y después de determinado tratamiento, cuyo valor central se espera que sea cero. Las diferencias iguales a cero son eliminadas y el valor absoluto de las desviaciones con respecto al valor central se ordena de menor a mayor. A aquellos valores idénticos se les asigna lugar medio dentro de la serie. La suma de los rangos se hace por separado para los signos positivos y los negativos. En este caso se emplea  $S$  para representar el valor menor de dichas sumas. Posteriormente,  $S$  deberá ser comparado con valores existentes en tablas estadísticas para determinar la aceptación o rechazo de la hipótesis nula, siempre de acuerdo al nivel de significancia elegido.

A continuación, se describe un ejemplo de la aplicación de una prueba de rangos con signo de Wilcoxon [29]:

El fabricante de un club de golf cree que el puntaje de los golfistas puede disminuir al usar su nuevo diseño de palo de golf. Por ello, se recabó el puntaje obtenido por 10 golfistas usando palos con el diseño anterior y el nuevo diseño. Con  $\alpha = 0.05$ , ¿Existe manera de

respaldar la afirmación del fabricante? Los puntajes obtenidos por los golfistas se muestran en la siguiente tabla:

Golfista	1	2	3	4	5	6	7	8	9	10
Puntaje (diseño antiguo)	89	84	96	74	91	85	95	82	92	81
Puntaje (nuevo diseño)	83	83	92	76	91	80	87	85	90	77

Con base en la afirmación, puede suponerse que:

- $H_0$ : El nuevo diseño no afecta el puntaje.
- $H_a$ : El nuevo diseño afecta el puntaje.

Esta prueba de rangos con signo de Wilcoxon es de una cola con  $\alpha = 0.05$  y, dado que la diferencia entre un par de datos es 0,  $n = 9$ . De acuerdo con la información extraída desde las tablas, el valor crítico es 8. Para encontrar el valor de la estadística de Wilcoxon es necesario efectuar los pasos antes descritos sobre la información de las muestras, obteniendo una tabla como la que se muestra a continuación:

Puntaje (diseño antiguo)	Puntaje (nuevo diseño)	Diferencia	Valor absoluto	Rangos	Rangos con signo
89	83	6	6	8	8
84	83	1	1	1	1
96	92	4	4	5.5	5.5
74	76	-2	2	2.5	-2.5
91	91	0	0		
85	80	5	5	7	7
95	87	8	8	9	9
82	85	-3	3	4	-4
92	90	2	2	2.5	2.5
81	77	4	4	5.5	5.5

De dicha tabla es posible calcular la suma de rangos positivos y negativos, obteniendo:  $W^+ = 8 + 1 + 5.5 + 7 + 9 + 2.5 + 5.5 = 38.5$  y  $W^- = -2.5 - 4 = -6.5$ , respectivamente. El valor de la estadística estará determinado entonces por el menor de dichos valores, ya que  $|-6.5| < |38.5|$ ,  $W_S = 6.5$ .

Debido a que el valor obtenido de la prueba es inferior al valor crítico ( $6.5 < 8$ ), se debería desechar la hipótesis nula. Por lo tanto, puede afirmarse que existe suficiente evidencia con un nivel de significancia de 5% para respaldar la afirmación de que el puntaje de los golfistas puede verse afectado al utilizar los palos con el nuevo diseño.

## 2.10.2 Prueba de Friedman

La prueba de Friedman es la alternativa no paramétrica a la prueba ANOVA para medidas repetidas. Su cálculo consiste en el ordenamiento de los datos por bloques para posteriormente ser reemplazados por sus respectivos rangos. Debido a que basa su funcionamiento en el uso de rangos, la prueba de Friedman es especialmente útil para probar los efectos de determinado tratamiento cuando las observaciones se encuentran expresadas a manera de rangos [30]. Una prueba de Friedman se compone de los siguientes elementos:

- Un grupo, conformado por medidas tomadas en tres o más bloques de condiciones experimentales.
- Una variable dependiente.

Y se consideran las siguientes suposiciones:

- El grupo es una muestra aleatoria de la población.
- Las muestras no tienen distribución normal.

La aplicación de la prueba considera las siguientes hipótesis:

- $H_0$ : No existe diferencia significativa entre las condiciones de experimentación dadas o la distribución de probabilidad de todas las condiciones de experimentación es la misma (las medianas son iguales). Es decir:  $H_0: Me_1 = Me_2 = Me_3 = \dots = Me_k$ , donde  $Me$  es la mediana.
- $H_a$ : Al menos dos condiciones de experimentación difieren entre sí (muestran diferencia significativa).

El método a seguir para la aplicación de la prueba se describe a continuación [32]:

1. Considérense como  $k$  el número de experimentos (3,4,...) y como  $n$  el número de bloques.
2. Calcular los rangos de cada bloque.
3. Sumar los rangos de cada experimento por separado; denotando dichas sumatorias como  $T_1, T_2, \dots, T_k$ .
4. Calcular la estadística de Friedman ( $F_r$ ), cuya distribución es la de  $\chi^2$  y está dada por:

$$F_r = \frac{12}{nk(k+1)} (T_1^2 + T_2^2 + \dots + T_k^2) - 3n(k+1) \quad (19)$$

5. Obtener el valor de  $p$  (como si se tratara de  $\alpha$ ) desde la tabla de la distribución de  $\chi^2$  con cola por la derecha para  $k - 1$  grados de libertad.
6. Comparar el valor obtenido con el valor crítico para determinar si se respalda o desecha la hipótesis establecida.

A continuación, se describe un ejemplo de la aplicación de la prueba no paramétrica de Friedman [32]:

Suponga que un alergólogo aplica tres pruebas de punción cutánea (pruebas de alergia), dos de ellas con estímulos en competencia y una de control, en los antebrazos de ocho pacientes, aleatorizando el orden de proximal a distal. Después de 15 minutos, el alergólogo asigna rangos a las reacciones de ronchas y eritema de cada paciente de acuerdo a su severidad (1= más severa) y se almacenaron en una tabla. Con base en ellos ¿Existe diferencia entre los resultados obtenidos? La tabla de resultados se muestra a continuación:

Paciente	Prueba 1	Prueba 2	Control
1	1	2	3
2	2	1	3
3	1	3	2
4	3	2	1
5	1	2	3
6	1	2	3
7	2	1	3
8	1	2	3

Para aplicar la prueba es necesario contabilizar los rangos e identificar determinados elementos, para ello es posible generar la siguiente tabla auxiliar:

Paciente	Prueba 1	Prueba 2	Control
1	1	2	3
2	2	1	3
3	1	3	2
4	3	2	1
5	1	2	3
6	1	2	3
7	2	1	3
8	1	2	3
<b>Suma de rangos</b>	<b>12</b>	<b>15</b>	<b>21</b>

El número de experimentos es 3, por lo que  $k=3$  y el número de bloques es 8, de manera que  $n=8$ . De la tabla anterior es posible extraer los valores de la suma de los rangos para cada experimento. Con dichos valores es posible calcular la estadística empleando la ecuación (19) de la siguiente manera:

$$F_r = \frac{12}{8(3)(4)}(144 + 225 + 441) - 3(8)(4) = 5.25 \quad (20)$$

Con base en lo observado en la tabla de la estadística  $\chi^2$ , el valor crítico para  $\alpha = .05$  con  $k - 1 = 2$  grados de libertad es 5.99. Ya que  $F_r < 5.99$ , el alergólogo no tiene evidencia suficiente para inferir que existe alguna diferencia en la eficacia de las pruebas o que las pruebas son más eficaces que la prueba de control.

# Capítulo 3

## Estado del arte

Como parte de la redacción de este trabajo, se realizó la revisión de una gran cantidad de trabajos que abordan diversas variantes del problema multiobjetivo de selección de cartera de proyectos. En este capítulo se describen de manera general los más relevantes y se contrastan con el trabajo propuesto en este trabajo.

### 3.1 Solución de problemas de cartera de proyectos públicos a partir de información del ranking de prioridades

En esta tesis doctoral [33], se aborda la versión estática del problema multiobjetivo de optimización de carteras de proyectos sociales empleado un modelo de discrepancia. En ella, se proponen algoritmos basados en la filosofía de colonias de hormigas, cada uno de ellos incorporando un elemento a la versión anterior. Los algoritmos propuestos son ACO-SPRI, que cuenta con un sistema de *ranking* para las prioridades, ACO-SOP que considera la incorporación de preferencias a través de clasificación y, finalmente una versión de ACO-SOP que incorpora la sinergia.

La factibilidad de los algoritmos propuestos fue validada mediante la resolución de tres instancias del estado del arte de 25, 40 y 100 proyectos, respectivamente, considerando como métrica de evaluación el número de soluciones no dominadas generadas.

Las principales contribuciones del trabajo son metodologías de bajo costo computacional que brindan soluciones con una calidad razonable, un modelo que permite evaluar carteras empleando un mecanismo de *ranking* de prioridades y tres algoritmos de resolución.

### **3.2 Artificial Bee Colony Algorithm Based on Novel Mechanism for Fuzzy Portfolio Selection**

En este artículo [34], se aborda el problema de selección de carteras de activos tomando en cuenta la incertidumbre presente en las tasas de retorno. En él, se propone la incorporación de dos nuevos mecanismos de aprendizaje a un algoritmo basado en la filosofía de colonias de abejas artificiales (*Artificial Bee Colony*, ABC por sus siglas en inglés). Los mecanismos de aprendizaje propuestos, denominados aprendizaje de direcciones y aprendizaje elitista (*direction learning* y *elite learning*), trabajan en conjunto en un mecanismo denominado *LL-mechanism*, que permite mejorar el desempeño de los algoritmos.

Para probar el mecanismo *LL-mechanism*, propone su incorporación a tres distintas variantes de algoritmos ABC del estado del arte ABC, GABC y CABC, generando los algoritmos LL-ABC, LL-GABC y LL-CABC respectivamente.

Los algoritmos propuestos fueron evaluados y comparados con sus versiones clásicas usando un conjunto de funciones *benchmark*. En la gran mayoría de los casos, los algoritmos propuestos obtuvieron mejores resultados que las versiones clásicas.

La principal contribución de este trabajo es el mecanismo denominado *LL-mechanism*, que permite mejorar el desempeño de algoritmos ABC. Además, los mecanismos de aprendizaje propuestos para ABC pueden ser de gran utilidad para la formulación de nuevos mecanismos de mejora.

### **3.3 Fuzzy portfolio optimization model under real constraints**

En este artículo [10], se aborda el problema de multiobjetivo de selección de carteras de activos en un ambiente difuso, modelando la incertidumbre presente en las tasas de retorno y diversos tipos de riesgo como conjuntos difusos. Se basa en la teoría de las posibilidades para modelar el retorno y la liquidez como medias probabilísticas y los riesgos de mercado y de liquidez como semivarianzas de baja probabilidad. En él, se proponen dos mecanismos para modelar el problema usando enfoques probabilísticos de media-semivarianza con restricciones reales.

La estrategia de resolución del modelo propuesto incluye un mecanismo que permite reducir el problema a problemas monoobjetivo y su posterior resolución mediante un algoritmo genético basado en representaciones numéricas, cuya factibilidad fue validada mediante la comparación con un algoritmo genético del estado del arte usando datos de la bolsa de valores de Shanghai.

La principal contribución de este trabajo radica en la formulación del problema y el mecanismo para reducir el problema a problemas monoobjetivo.

### **3.4 A new bi-objective fuzzy portfolio selection model and its solution through evolutionary algorithms**

En este artículo [35], se aborda la versión biobjetivo del problema de selección de carteras de activos. El modelo propuesto cuenta con dos objetivos de maximización: *Sharpe ratio* (SR) y *Value at Risk ratio* (VR). Los parámetros relacionados a dichos objetivos son representados mediante números difusos triangulares.

La estrategia de resolución para el modelo propuesto se basa en algoritmos genéticos multiobjetivo. Los algoritmos usados para su resolución son del tipo MOCeLL (*MultiObjective Cellular genetic algorithm*), AbySS (*Archive-based hybrid Scatter Search*) y NSGA-II (*Non-dominated Sorting Genetic Algorithm II*).

Los algoritmos propuestos fueron evaluados en la resolución de un conjunto de datos de la bolsa de valores de Shenzhen y sus desempeños fueron comparados usando métricas clásicas de evaluación de algoritmos multiobjetivo.

### **3.5 Fuzzy Portfolio Optimization Model with Fuzzy Numbers**

Este artículo [36] aborda una versión mono objetivo del problema de selección de carteras de activos. Dada la importancia del riesgo presente en las tareas de inversión en bolsas de valores, el autor propone un nuevo modelo matemático para el problema que tome en cuenta la incertidumbre presente en la percepción del riesgo y en las posibles variaciones del retorno.

En este trabajo, se propone modelar el problema como un problema de maximización en donde la función objetivo representa el retorno esperado y deberá estar restringida, ya que se desea que el riesgo no sea mayor a un valor de tolerancia previamente definido, en donde el riesgo y el retorno esperado son modelados como números difusos triangulares.

Como estrategia de resolución, se propone emplear el criterio de probabilidad basado en dominancia para convertir el problema en un problema de programación lineal clásica.

La principal aportación de este trabajo es el mecanismo empleado transformar el modelo difuso en un problema de programación lineal.

### **3.6 Modeling and Optimizing the Multi-Objective Portfolio Optimization Problem with Trapezoidal Fuzzy Parameters**

En este artículo [3] se modela la incertidumbre presente en los parámetros de la variante multiobjetivo del problema de selección de carteras de proyectos. En el modelo propuesto, todos los parámetros (costos y presupuestos) se representan mediante números difusos trapezoidales. Además, se propone la incorporación de mecanismos propios de los números difusos a algoritmos del estado del arte.

En este trabajo, se proponen dos algoritmos basados en la metodología NSGA-II, T-NSGA-II-CD y T-NSGA-II-SSD, éste último incorporando el estimador de densidad SSD (usado normalmente en el algoritmo FAME) al modelo clásico de NSGA-II para brindarle una mejor distribución de soluciones. También se propone un algoritmo basado en el algoritmo FAME, denominado T-FAME, siendo éste una adaptación del algoritmo clásico para funcionar con parámetros difusos trapezoidales.

El desempeño de los algoritmos propuestos en este trabajo fue evaluado usando un conjunto de instancias y métricas clásicas. Los resultados de la experimentación muestran que el algoritmo con mejor desempeño fue T-FAME.

Las principales contribuciones de este trabajo son el modelo del problema de selección de cartera de proyectos que aborda la incertidumbre de sus parámetros mediante números difusos trapezoidales, las dos adaptaciones del algoritmo NSGA-II: T-NSGA-II-CD y T-NSGA-II-SSD, y el algoritmo T-FAME.

### **3.7 Modelando la imprecisión del problema de cartera de proyectos con filosofía gris**

Esta tesis [2], aborda la variante multiobjetivo de optimización de carteras de proyectos en su versión estática y se enfoca en modelar la incertidumbre presente en sus parámetros usando la metodología de números de tipo intervalo. En ella trabajo se propone una adaptación del algoritmo clásico NSGA-II para funcionar con números en intervalo denominado I-NSGA-II.

En este trabajo se toma en cuenta la incorporación de preferencias del tomador de decisiones. Las preferencias son incorporadas en un mecanismo *a priori* empleando relaciones de clasificación.

El desempeño del algoritmo propuesto fue evaluado mediante la cantidad de soluciones no dominadas de los frentes obtenidos luego de resolver un conjunto de instancias del estado del arte de distintos tamaños. El algoritmo propuesto mostró ser competitivo con los mejores algoritmos del estado del arte que abordan esta variante del problema al compararse con los algoritmos I-NOSGA y A<sup>2</sup> NSGA-II. Además, se estudió el efecto de la variación de los valores de los parámetros en las relaciones de clasificación.

Como principales aportaciones de este trabajo de tesis pueden considerarse las siguientes: un nuevo modelo de la incertidumbre presente en los parámetros del problema de selección de cartera de proyectos usando números en intervalos, una adaptación del algoritmo NSGA-II que puede operar con parámetros en intervalos y un nuevo modelo de incorporación de preferencias.

**Tabla 1.** Tabla comparativa de trabajos relacionados.

<b>Trabajo</b>	<b>Autor(es)</b>	<b>Problema</b>	<b>Algoritmo(s)</b>	<b>Enfoque</b>	<b>Parámetros</b>
Solución de problemas de cartera de proyectos públicos a partir de información del ranking de prioridades (2017) [33]	Bastiani Medina, S	Selección de cartera de proyectos públicos	ACO-SPRI, ACO-SOP	Incorporación de preferencias	Reales
Artificial Bee Colony Algorithm Based on Novel Mechanism for Fuzzy Portfolio Selection (2019) [34]	Gao, Weifeng, et al	Selección de cartera activos	LL-ABC, LL-GABC, LL-CABC	Modelar la incertidumbre del retorno	Números difusos
Fuzzy portfolio optimization model under real constraints (2011) [10]	Liu, Yong-Jun, y Wei-Guo Zhang	Selección de cartera activos	Algoritmo genético	Modelar riesgo, reducir a monoobjetivo	Números difusos
A new bi-objective fuzzy portfolio selection model and its solution through evolutionary algorithms (2019) [35]	Kar, Mohuya B., et al	Selección de cartera activos	MOCcell, AbySS, NSGA-II	Modelar riesgo	Números difusos triangulares
Fuzzy Portfolio Optimization Model with Fuzzy Numbers (2011) [36]	Li, Chunquan, y Jianhua Jin	Selección de cartera activos	Programación lineal	Monoobjetivo, modelar riesgo	Números difusos triangulares
Modeling and Optimizing the Multi-Objective Portfolio Optimization Problem with Trapezoidal Fuzzy Parameters (2021) [3]	Estrada-Padilla, Alejandro, et al	Selección de cartera proyectos	T-NSGA-II-CD, T-NSGA-II-SSD, T-FAME	Modelar funciones objetivo y restricciones	Números difusos trapezoidales
Modelando la imprecisión del problema de cartera de proyectos con filosofía gris (2018) [2]	Balderas Jaramillo, Fausto	Selección de cartera de proyectos	I-NSGA-II	Incorporación de preferencias	Intervalos
Este trabajo	Velasco Ocejo, Luis M.	Selección de cartera de proyectos	Tri-NSGA-II-CD Tri-NSGA-II-SSD	Modelar funciones objetivo y restricciones	Números difusos triangulares

Como puede observarse en la Tabla 1, la mayoría de los trabajos más relevantes encontrados en la literatura se centra en modelar algunas formas de la incertidumbre presente en el proceso de inversión en carteras de proyectos, en el diseño de estrategias que permitan incorporar preferencias a los algoritmos con el propósito de facilitar la tarea del tomador de decisiones o en maneras de manipular el problema para facilitar su resolución, ya sea reduciéndolo a varios problemas mono objetivo o transformándolo en un problema de programación lineal entera.

En este trabajo de tesis propone una estrategia que permite modelar e incorporar a algoritmos de solución la incertidumbre del problema de cartera de proyectos mediante números difusos de tipo triangular, estrategia comúnmente utilizada para modelar expresiones en lenguaje natural del tipo “*cercano a*” [5]. Brindando a la entidad encargada del análisis del problema contar con otra alternativa durante el proceso de la selección de la estrategia más apropiada para la forma en que se presenta la incertidumbre.

# Capítulo 4

## Modelando el Problema de Selección de Cartera de Proyectos con Parámetros Triangulares

En este capítulo se describe el modelo matemático propuesto para incorporar la incertidumbre al problema de selección de cartera de proyectos mediante números difusos triangulares. Además, se detalla la el procedimiento necesario para la evaluación de factibilidad de las soluciones candidata.

### 4.1 Modelo matemático

El modelo propuesto consta de los siguientes elementos:

Sean  $n$  el número de proyectos a considerar,  $\widetilde{C}$  el presupuesto total disponible,  $O$  el número de objetivos,  $\widetilde{c}_i$  el costo del proyecto  $i$ ,  $\widetilde{b}_{ij}$  el beneficio obtenido de la realización del proyecto  $i$  en el objetivo  $j$ ,  $K$  y  $M$  el número de áreas y regiones a considerar respectivamente,  $\widetilde{A}_k^{\min}$  y  $\widetilde{A}_k^{\max}$  los límites inferior y superior del presupuesto asignado al área  $k$ ,  $\widetilde{R}_m^{\min}$  y  $\widetilde{R}_m^{\max}$  los límites inferior y superior del presupuesto asignado a la región  $m$ . Las variables  $a_i, b_i$  son valores que representan el área y la región a las que pertenece el proyecto  $i$ .  $x = (x_1, x_2, \dots, x_n)$  es un vector binario que especifica los proyectos incluidos en un determinado portafolio. Si  $x_i = 1$ , el proyecto  $i$  se encuentra dentro del portafolio, en caso contrario, el proyecto  $i$  no está incluido en él. Con los elementos antes mencionados, es posible modelar el problema de la siguiente manera:

$$\text{Max } Z = (\widetilde{z}_1, \widetilde{z}_2, \dots, \widetilde{z}_n) \quad (21)$$

Donde:

$$\tilde{z}_j = \sum_{i=1}^n \tilde{b}_{ij} x_i; \quad j = 1, 2, \dots, O \quad (22)$$

Sujeto a:

$$\sum_{i=1}^n \tilde{c}_i x_i \leq \tilde{C} \quad (23)$$

$$\tilde{A}_k^{min} \leq \sum_{i=1, a_i=k}^n \tilde{c}_i x_i \leq \tilde{A}_k^{max}; \quad k = 1, 2, \dots, K \quad (24)$$

$$\tilde{R}_m^{min} \leq \sum_{i=1, b_i=m}^n \tilde{c}_i x_i \leq \tilde{R}_m^{max}; \quad m = 1, 2, \dots, M \quad (25)$$

$$x_i \in \{0, 1\}, i = 1, 2, \dots, n \quad (26)$$

En dicha formulación, todas aquellas variables con tilde (~) representan números difusos triangulares.

La función objetivo maximiza las contribuciones a cada uno de los objetivos (21). El valor de los objetivos se calcula mediante la adición de la contribución que cada uno de los proyectos presentes en el portafolio representa para cada objetivo (22). La restricción de presupuesto (23) asegura que la suma de los costos de inversión de los proyectos seleccionados no exceda el presupuesto total. De manera similar, las restricciones de área (24) y región (25) aseguran que la suma de los costos de los proyectos pertenecientes a determinada área y región se encuentren dentro de los límites de presupuesto establecidos. Las soluciones al problema serán vectores binarios de tamaño  $n$  que representarán los proyectos incluidos dentro del portafolio (26).

## 4.2 Representación de soluciones y validación de factibilidad

Como se mencionó en la formulación del problema (ecuación 26), las soluciones pueden representarse mediante un vector binario, cuyos valores representan la inclusión o ausencia de cada uno de los proyectos en el portafolio en cuestión, por lo que el espacio de posibles soluciones para un problema de  $n$  proyectos es  $S = \{0, 1\}^n$ .

1	1	0	0	1	0	1	1
---	---	---	---	---	---	---	---

**Figura 9.** Ejemplo de representación de una cartera.

En la Figura 9 puede observarse un ejemplo de solución candidata para determinada instancia del problema. Una solución candidata deber ser evaluada para determinar un vector de objetivos difusos y su factibilidad se debe verificar para comprobar que satisface las restricciones del modelo matemático del problema. Los parámetros difusos triangulares son utilizados en ambos procesos y, para compararlos, será necesario el cálculo de GMI de cada uno de ellos.

Para verificar la factibilidad de una solución, es necesario aplicar GMI sobre el modelo matemático mencionado anteriormente (sección 4.1), como se muestra a continuación:

$$Max Z = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n) \quad (27)$$

Donde:

$$\tilde{z}_j = P \left( \sum_{i=1}^n \tilde{b}_{ij} x_i \right); \quad j = 1, 2, \dots, O \quad (28)$$

Sujeto a:

$$P \left( \sum_{i=1}^n \tilde{c}_i x_i \right) \leq P(\tilde{C}) \quad (29)$$

$$P(\tilde{A}_k^{min}) \leq P \left( \sum_{i=1, a_i=k}^n \tilde{c}_i x_i \right) \leq P(\tilde{A}_k^{max}); \quad k = 1, 2, \dots, K \quad (30)$$

$$P(\tilde{R}_m^{min}) \leq P \left( \sum_{i=1, b_i=m}^n \tilde{c}_i x_i \right) \leq P(\tilde{R}_m^{max}); \quad m = 1, 2, \dots, M \quad (31)$$

Bajo este mecanismo de evaluación, si se considera la siguiente instancia simplificada (sin tomar en cuenta áreas ni regiones) con  $n = 4$ ,  $\tilde{C} = (19000, 1868, 19320)$  y  $O = 2$  con los siguientes proyectos:

Proyecto	$\tilde{c}_i$	$\tilde{b}_{i,0}$	$\tilde{b}_{i,1}$
0	(9695, 9308, 10082)	(7960, 7642, 8278)	(240, 231, 249)
1	(8635, 8290, 8980)	(8860, 8506, 9214)	(420, 404, 436)
2	(6140, 5895, 6385)	(3990, 3831, 4149)	(115, 111, 119)
3	(9430, 9053, 9807)	(4070, 3908, 4232)	(415, 399, 431)

Si se forma la cartera A:

1	0	1	0
---	---	---	---

La evaluación de su función objetivo estará dada por:

$$\text{Maximizar } Z = (\tilde{z}_1, \tilde{z}_2) \quad (32)$$

Donde:

$$\tilde{z}_1 = (7960, 7642, 8278) + (3990, 3831, 4149) = (11950, 11473, 12427) \quad (33)$$

$$\tilde{z}_2 = (240, 231, 249) + (115, 111, 119) = (355, 342, 368) \quad (34)$$

Y su costo de inversión:

$$\sum_{i=1}^n \tilde{c}_i x_i = (9695, 9308, 10082) + (6140, 5895, 6385) = (15835, 15203, 16467) \quad (35)$$

Para verificar la factibilidad de la solución es necesario calcular

$$P\left(\sum_{i=1}^n \tilde{c}_i x_i\right) = P(15835, 15203, 16467) = 16045.66 \quad (36)$$

y

$$P(\tilde{C}) = P(19000, 18680, 19320) = 19106.66 \quad (37)$$

Con dichos valores, es posible emplear la ecuación (28) y verificar el cumplimiento de las condiciones de factibilidad:

$$16045.66 \leq 19106.66 \therefore \text{La cartera A es factible} \quad (38)$$

# Capítulo 5

## Algoritmo propuesto: Tri-NSGA-II

**E**n esta sección se describe el diseño del algoritmo principal propuesto en el presente trabajo: Tri-NSGA-II. Este algoritmo es una adaptación del algoritmo genético basado en no dominancia clásico NSGA-II [37] para trabajar con parámetros difusos triangulares. Al igual que en el algoritmo original, Tri-NSGA-II presenta un enfoque generacional para abordar el problema de selección de portafolio. El algoritmo requiere como entrada una instancia del problema que contiene los valores difusos de cada uno de los parámetros mencionados en la sección anterior y su salida consiste en una aproximación al frente de Pareto de dicha instancia.

Al tratarse de una adaptación, Tri-NSGA-II se compone de los mismos elementos que el algoritmo original, incluyendo el cálculo del estimador de densidad *Crowding Distance* (CD, por sus siglas en inglés). Sin embargo, Tri-NSGA-II incorpora a su funcionamiento otro estimador de densidad, conocido como *Spatial Spread Deviation* (SSD, por sus siglas en inglés), que permite una mejor distribución de las soluciones de cada uno de los frentes y obtener mejores resultados en problemas con muchos objetivos [38]. Con el propósito de facilitar la distinción entre ambas implementaciones del algoritmo Tri-NSGA-II, se utilizará la siguiente nomenclatura:

- a) Tri-NSGA-II-CD: Implementación del algoritmo NSGA-II con parámetros difusos triangulares y *Crowding Distance* (CD) como estimador de densidad.
- b) Tri-NSGA-II-SSD: Implementación del algoritmo NSGA-II con parámetros difusos triangulares y *Spatial Spread Deviation* (SSD) como estimador de densidad.

El pseudocódigo del algoritmo Tri-NSGA-II se muestra a continuación.

---

**Algoritmo 1** Tri-NSGA-II

---

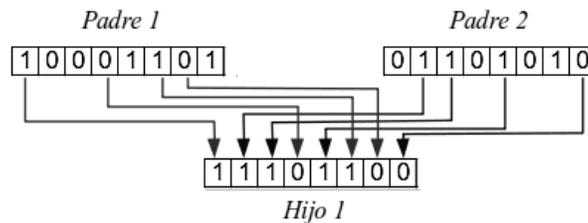
**Entrada:** Instancia del problema de selección de cartera de proyectos con parámetros difusos triangulares

**Salida:** Frente de Pareto aproximado

- 1:  $pop \leftarrow$  Generar población inicial
  - 2: Evaluar todas las soluciones de  $pop$
  - 3: Ordenar  $pop$  por **no dominancia**  $\triangleright$  *Non-dominated Sorting*
  - 4: Calcular los valores de densidad de las soluciones de  $pop$   $\triangleright$  *(CD/SSD)*
  - 5: Ordenar  $pop$  por los valores de densidad
  - 6: **for**  $i = 0$  to  $numIteraciones$  **do**  $\triangleright$  Ciclo principal
  - 7:    $popc \leftarrow$  Generar hijos (sólo factibles)  $\triangleright$  *Cruza*
  - 8:    $popm \leftarrow$  Generar soluciones mutadas (sólo factibles)  $\triangleright$  *Mutación*
  - 9:    $qt \leftarrow cpop \cup mpop$
  - 10:    $pop \leftarrow pop \cup qt$   $\triangleright$  *Actualizar población*
  - 11:   Evaluar todas las soluciones de  $pop$
  - 12:   *Non-dominated Sorting*( $pop$ )
  - 13:   Calcular los valores de densidad de las soluciones de  $pop$
  - 14:   Ordenar  $pop$  por los valores de densidad
  - 15:   Cortar  $pop$  al tamaño original
  - 16:   *Non-dominated Sorting*( $pop$ )
  - 17:   Calcular los valores de densidad de las soluciones de  $pop$
  - 18:   Ordenar  $pop$  por los valores de densidad
  - 19: **end for**
  - 20: **return**  $front(0)$   $\triangleright$  *Frente de Pareto aproximado*
- 

## 5.1 Operador de cruza: Cruza Uniforme

El operador de cruza usado en el diseño de Tri-NSGA-II es un operador clásico ampliamente utilizado en algoritmos genéticos. En él, se seleccionan dos soluciones de la población (padres) para generar descendencia y los genes de cada una de ellas tienen la misma probabilidad de ser transferidos a su descendencia [39] [40].



**Figura 10.** Ejemplo de cruza uniforme.

## 5.2 Operador de mutación: *One bit flip*

Es un método clásico de mutación de soluciones con representación binaria. Consiste en seleccionar de manera aleatoria una posición de una solución y cambiar su valor binario (de 0 a 1 o viceversa).

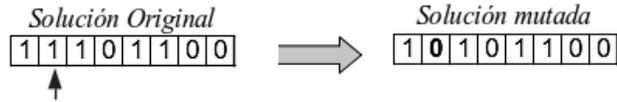


Figura 11. Ejemplo de mutación con *one bit flip*.

## 5.3 Generación de la población inicial

La población inicial del algoritmo Tri-NSGA-II comprende un número previamente definido de portafolios, generados de manera aleatoria. Cada solución generada es evaluada para verificar su factibilidad. La población inicial deberá estar conformada únicamente por soluciones factibles.

## 5.4 Ordenamiento de la población

El ordenamiento de la población constituye una parte fundamental en el funcionamiento del algoritmo Tri-NSGA-II-SSD. En el algoritmo clásico NSGA-II, las soluciones que conforman la población son ordenadas en dos etapas: la etapa elitista, denominada *Non-dominated Sorting* que asegura la permanencia de las mejores soluciones, y la etapa de diversificación, conocida como *Crowding Distance Sorting* (CD, por sus siglas en inglés), que busca asegurar que las soluciones sean suficientemente distintas entre sí. En el algoritmo propuesto en este trabajo de tesis, se incorporó el estimador de densidad *Spatial Spread Deviation* reemplazando al estimador *Crowding Distance* en la etapa de diversificación.

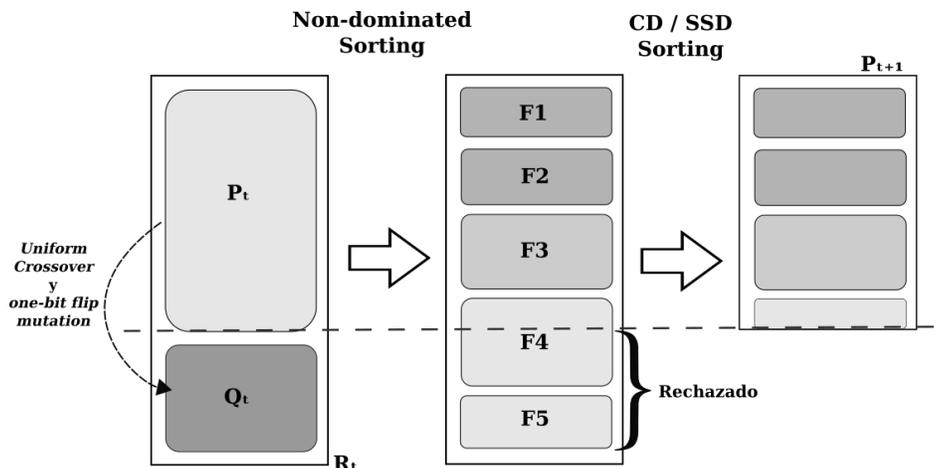


Figura 12. Diagrama general de Tri-NSGA-II.

La Figura 12 permite observar claramente las etapas del algoritmo propuesto y en cuál de ellas se aplican los operadores genéticos.

### 5.4.1 *Non-dominated Sorting*

El ordenamiento por no dominancia es un método de ordenamiento que tiene como propósito preservar la calidad de los portafolios presentes en la población, su funcionamiento puede dividirse en dos etapas, la primera de ellas consiste en la generación de un grupo de soluciones, denominado “frente de soluciones no dominadas”. Para determinar las soluciones que integran dicho frente, los vectores de objetivos de cada una de las soluciones de la población serán comparados entre sí para verificar la relación de dominancia que existe entre ellas. Aquellas soluciones que no sean dominadas por ninguna otra solución formarán parte del frente.

La segunda etapa del proceso de ordenamiento consiste en la generación de frentes subsecuentes con el conjunto de portafolios restante. Para ello, cada uno de los portafolios presentes en el frente recién creado es explorado y el contador de dominancia de cada uno de los portafolios dominados por ellos se decrementa en 1, en caso de que el contador de alguno de ellos llegue a 0, será agregado al nuevo frente.

### 5.4.2 *Crowding Distance Sorting*

El valor de *Crowding Distance* (o distancia de amontonamiento) de una solución es un valor estimado de la densidad de soluciones que la rodean [37].

Este tipo de ordenamiento consiste en explorar cada uno de los frentes generados en la etapa anterior, calculando el valor de la métrica CD para cada uno de los portafolios contenidos en ellos. Posteriormente, los portafolios de cada frente serán reorganizados de acuerdo a sus respectivos valores de CD. Para calcular el valor de la métrica es necesario ordenar las soluciones de cada frente por sus valores objetivo, posteriormente, se deberá asignar  $\infty$  como valor de CD de la mayor y menor solución respectivamente. El valor de CD del resto de las soluciones puede calcularse mediante la fórmula:

$$d_{Ij}m = d_{Ij}m + \frac{f_{m^{j+1}}^{I^m} - f_{m^{j-1}}^{I^m}}{f_m^{max} - f_m^{min}} \quad (39)$$

Donde  $d$  es el valor de la métrica,  $I$  es la posición de la solución dentro de la población,  $j$  es la posición de la solución con respecto a las demás soluciones del frente  $f$  es el valor objetivo,  $m$  es el objetivo actual. El proceso se realiza con cada uno de los  $m$  objetivos, la sumatoria de dichos valores representará el valor de CD de dicha solución. Una vez calculada la métrica para cada uno de los portafolios, el frente deberá ser ordenado en forma decreciente.

En el presente trabajo, dado que se propone la incorporación del estimador de densidad SSD, se realizó también la implementación de una variante del algoritmo con el estimador CD (denominada Tri-NSGA-II-CD) empleado en el algoritmo NSGA-II original con el propósito de contar con un punto de comparación de los resultados obtenidos por el algoritmo Tri-NSGA-II-SSD.

### 5.4.3 Spatial Spread Deviation Sorting

La Desviación de Dispersión Espacial (SSD) es un estimador de densidad cuyo propósito de incorporación es evitar que el espacio que separa a las soluciones presentes en cada uno de los frentes sea muy grande. Además, se ha comprobado que la diversidad de soluciones del frente aproximado obtenido por un algoritmo NSGA-II que incorpora SSD se incrementa significativamente en comparación con otros estimadores de densidad [38].

De manera similar al ordenamiento por CD, los valores de los objetivos de las soluciones de cada frente son normalizadas para posteriormente calcular su respectivo valor de SSD con las fórmulas siguientes:

$$temp1(i) = \frac{1}{n-1} \sqrt{\sum_{j=1}^n (D(i,j) - (D_{max} - D_{min}))^2} \quad \forall i \neq j \quad (40)$$

$$temp2(i) = \sum_{j \in K} \frac{D_{max} - D_{min}}{D(i,j)} \quad (41)$$

$$SSD(i) = SSD_0(i) + temp1(i) + temp2(i) \quad (42)$$

Donde  $D(i,j)$  es la distancia de la solución  $i$  a la solución  $j$ .  $D_{max}$  y  $D_{min}$  son la mayor y menor distancia entre las soluciones, respectivamente.  $K$  es el número de  $k$  vecinos más cercanos a la solución  $i$  y  $SSD_0$  es el valor inicial de SSD de la solución  $i$ , el cual, en caso de ser uno de los valores extremos del frente de objetivos normalizados deberá tomar el valor de  $-\infty$ .

# Capítulo 6

## Resultados experimentales

Con el propósito de evaluar el funcionamiento y desempeño del algoritmo Tri-NSGA-II, propuesto en el presente trabajo de tesis, se realizaron experimentos para evaluar su funcionamiento y el impacto de la incorporación del estimador de densidad SSD en comparación con la versión clásica que hace uso de CD.

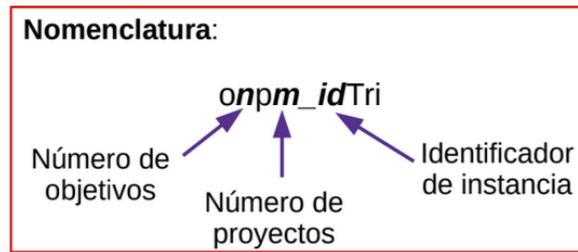
### 6.1 Instancias del problema

Para realizar la experimentación, se seleccionó un grupo de instancias de distintos tamaños y número de objetivos. Las instancias utilizadas fueron originalmente diseñadas para el problema de selección de portafolio con parámetros del tipo intervalo [11] por lo que fueron procesadas con el propósito de convertir sus parámetros a números difusos triangulares. La estructura de las instancias es la siguiente:

- Presupuesto total.
- Número de objetivos.
- Número de áreas.
- Límites de presupuesto para cada área (uno por línea).
- Número de regiones.
- Límites de presupuesto para cada región (uno por línea).
- Número de proyectos.
- Descripción de proyectos (uno por línea), incluye: costo, área, región, aportación a objetivos.

Las instancias resueltas durante la experimentación se identifican mediante los siguientes nombres:

- o2p20\_Tri
- o4p25\_0Tri
- o4p25\_1Tri
- o4p25\_2Tri
- o4p25\_3Tri
- o4p25\_4Tri
- o4p25\_5Tri
- o9p100\_0Tri
- o9p100\_1Tri
- o9p100\_2Tri



**Figura 13.** Nomenclatura de instancias.

Como puede observarse en la Figura 13, el tamaño de las instancias (número de objetivos y número de proyectos) está indicado mediante la nomenclatura que se les asignó.

```

80000 76800 83200
2
20 80
11000 550
1500 75
3
14810 13060 16560 47995 46245 49745
14810 13810 15810 47995 47895 48095
14810 13210 16410 47995 46545 49445
2
23525 22775 24275 68000 67950 68050
23525 23325 23725 68000 67900 68100
20
9695 9308 10082 1 1 7960 7642 8278 240 231 249
8635 8290 8980 2 1 8860 8506 9214 420 404 436
6140 5895 6385 3 1 3990 3831 4149 115 111 119
9430 9053 9807 1 2 4070 3908 4232 415 399 431
6310 6058 6562 1 2 6000 5760 6240 435 418 452
8600 8256 8944 3 1 4835 4642 5028 400 384 416
9360 8986 9734 1 1 9415 9039 9791 305 293 317
7080 6797 7363 2 1 9530 9149 9911 360 346 374
8715 8367 9063 1 2 6530 6269 6791 130 125 135
8700 8352 9048 3 1 5445 5228 5662 375 360 390
7490 7191 7789 1 1 1305 1253 1357 325 312 338
5490 5271 5709 1 2 5480 5261 5699 300 288 312
5450 5232 5668 1 2 3825 3672 3978 110 106 114
9585 9202 9968 2 2 1225 1176 1274 475 456 494
5550 5328 5772 3 1 3515 3375 3655 115 111 119
9875 9480 10270 2 1 9085 8722 9448 350 336 364
9925 9528 10322 2 1 9675 9288 10062 330 317 343
6780 6509 7051 2 1 5835 5602 6068 310 298 322
9970 9572 10368 1 1 7305 7013 7597 490 471 509
5205 4997 5413 1 2 1675 1608 1742 485 466 504

```

**Figura 14.** Contenido de instancia o2p20\_Tri.

En la Figura 14, se muestra un ejemplo del contenido de una instancia para el problema de selección de cartera de proyectos con parámetros difusos triangulares.

### 6.1.1 Conversión de instancias

Como se ha mencionado anteriormente, los números difusos triangulares constan de 3 componentes, un límite inferior y uno superior que determinan su extensión y el punto en el que la función de membresía toma el valor de 1. Dado que los números en intervalos constan de un límite superior y un límite inferior, es posible convertirlo en un número difuso triangular asignándole un valor al punto máximo de la función de membresía. En el presente trabajo, dicho valor corresponde al punto medio del intervalo, por ejemplo:

Dado el presupuesto total de una instancia de problema en formato intervalo:

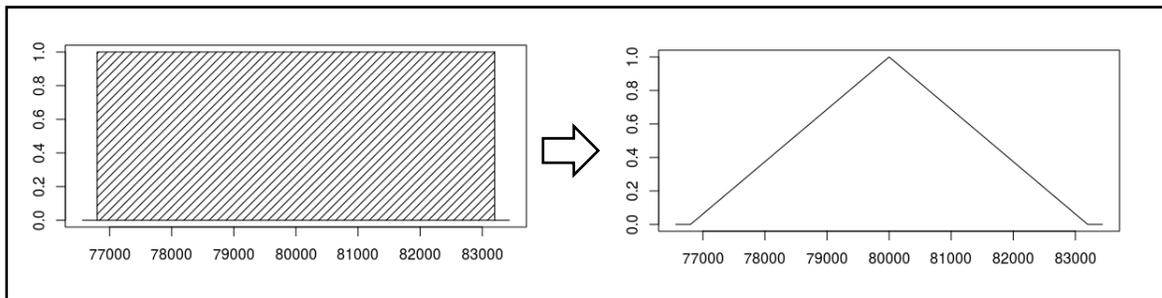
$$C = [76800, 83200] \quad (43)$$

Sea  $\tilde{C}$  la representación de  $C$  como número difuso triangular. Si se consideran los límites del presupuesto (43) como los valores de  $\alpha$  y  $\beta$  respectivamente de  $\tilde{C}$  y  $a$  se define como el punto medio del intervalo  $C$ , entonces es posible calcular  $a$  de la siguiente manera:

$$a = \frac{83200 + 76800}{2} = 80000 \quad (44)$$

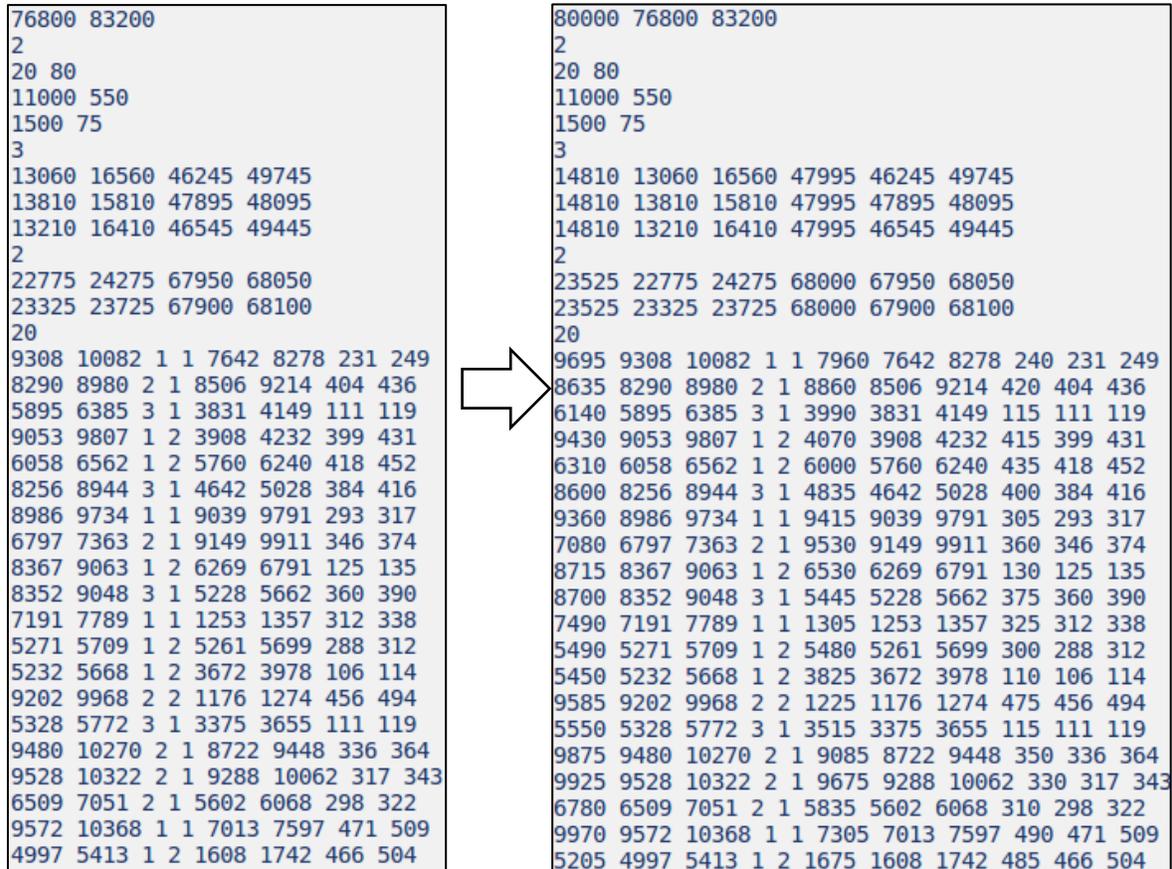
De manera que:

$$\tilde{C} = (80000, 76800, 83200) \quad (45)$$



**Figura 15.** Ejemplo de conversión de un número en intervalo a número difuso triangular.

Como se ilustra en la figura 15, la conversión no altera el rango de valores, únicamente le asigna una función de membresía triangular al intervalo. De esta manera, es posible convertir todos los parámetros de una instancia al dominio de los números difusos triangulares.



**Figura 16.** Conversión de la instancia *o2p20\_Gris* (intervalos, lado izquierdo) a la instancia *o2p20\_Tri* (números difusos triangulares, lado derecho).

## 6.2 Algoritmo exhaustivo

Para validar el funcionamiento del algoritmo propuesto, es necesario comparar los resultados obtenidos de su ejecución con el frente óptimo de la instancia en cuestión, lo que implica la obtención de su solución exacta. Sin embargo, debido a que al resolver de manera exhaustiva una instancia representa un costo computacional alto, se optó por resolver una única instancia de 2 objetivos, la instancia *o2p20\_Tri*.

El algoritmo empleado para resolver de manera exhaustiva las instancias se muestra a continuación:

---

**Algoritmo 2** BruteForceSolver

---

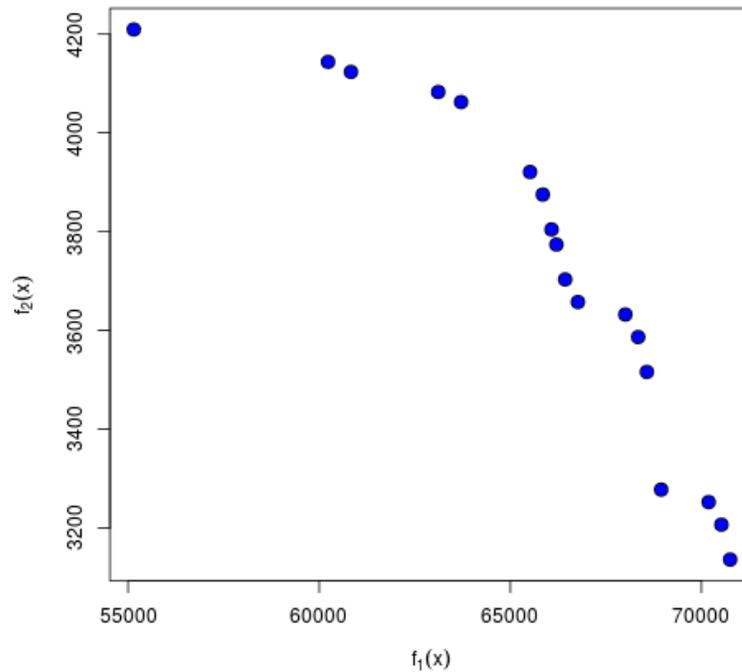
**Entrada:** Instancia del problema de selección de cartera de proyectos con parámetros triangulares

**Salida:** Frente de Pareto

```
1:  $instSize \leftarrow$  Tamaño de la instancia (número de proyectos)
2:  $portfolio \leftarrow$  Inicializar arreglo binario de tamaño  $instSize$ 
3:  $feasiblePorts \leftarrow$  Inicializar lista de soluciones factibles
4: for  $i = 0$  to  $2^{instSize} - 1$  do ▷  $2^n - 1$  combinaciones
5:    $j \leftarrow instSize - 1$ 
6:   while  $portfolio[j] \neq 0$  do
7:      $portfolio[j] \leftarrow 0$ 
8:      $j \leftarrow j - 1$ 
9:   end while
10:   $portfolio[j] \leftarrow 1$ 
11:  if  $isFeasible(portfolio)$  then ▷ Validar factibilidad
12:     $feasiblePorts \cup portfolio$ 
13:  end if
14: end for
15: Calcular No-dominancia entre las soluciones de  $feasiblePorts$ 
16:  $paretoFront \leftarrow$  Soluciones no dominadas de  $feasiblePorts$ 
17: return  $paretoFront$  ▷ Frente de Pareto
```

---

Para la instancia *o2p20\_Tri* se obtuvieron 1,048,576 ( $2^{20}$ ) soluciones, de las cuales 30681 resultaron factibles y únicamente 18 forman parte del frente de Pareto.



**Figura 17.** Representación gráfica del frente de Pareto de la instancia *o2p20\_Tri* (valores de GMI).

## 6.3 Condiciones de experimentación

### 6.3.1 Hardware y Software

Los experimentos descritos en el presente trabajo fueron realizados en un equipo con las siguientes características:

- Procesador: AMD A8-6410 *Quad Core* a 2.4 GHz.
- Memoria RAM: 6GB (1x4GB, 1x2GB) a 1600MHz.
- Entorno de desarrollo: Eclipse IDE.
- Lenguaje de programación: Java.
- Versión de JDK: 8.

### 6.3.2 Configuración de parámetros de algoritmos

La configuración de parámetros usada durante el proceso de experimentación se lista a continuación:

- Tamaño de la población: 50.
- Número de iteraciones: 100.
- Porcentaje de población a cruzar 20%.
- Porcentaje de población a mutar 50%.

## 6.4 Validación de funcionamiento

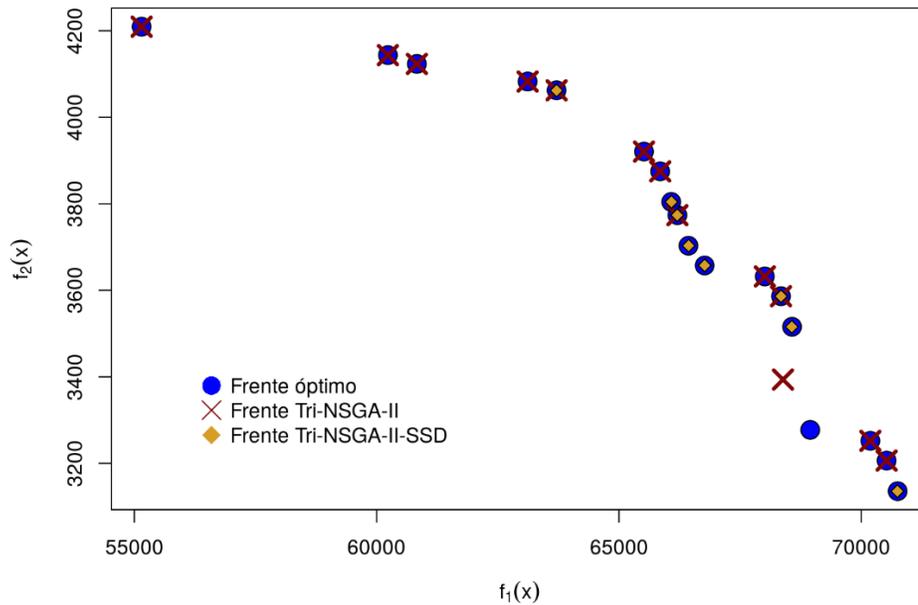
Como parte del proceso de experimentación, se desarrolló un algoritmo capaz de resolver de forma exhaustiva una instancia pequeña (instancia *o2p20\_Tri*) del problema en cuestión con el propósito de contar con el frente de Pareto (frente óptimo) de al menos una de las instancias.

El primero de los experimentos consistió en resolver la instancia *o2p20\_Tri* con los algoritmos Tri-NSGA-II-CD y Tri-NSGA-II-SSD. Posteriormente, se calculó el valor de GMI del vector de objetivos de cada una de las soluciones pertenecientes a los frentes aproximados obtenidos. De igual manera, se calcularon los valores de GMI del vector de objetivos de cada una de las soluciones que conforman el frente de Pareto de dicha instancia. Por último, los resultados obtenidos se compararon con los resultados del frente de Pareto para validar el correcto funcionamiento de los algoritmos.

A continuación, se muestra la comparación de los valores obtenidos por ambas variantes del algoritmo y el frente de Pareto para la mencionada instancia.

**Tabla 2.** Comparación de frentes para *o2p20\_Tri* (valores de GMI).

Frente óptimo		Tri-NSGA-II-CD		Tri-NSGA-II-SSD	
Objetivo 1	Objetivo 2	Objetivo 1	Objetivo 2	Objetivo 1	Objetivo 2
66767.6667	3657.33333	68343.3333	3586.33333	68343.3333	3586.33333
68571.3333	3515.66667	68009	3632	63712.3333	4062.33333
68946.3333	3277.33333	55149.6667	4209.33333	66205.3333	3773.66667
70750	3135.66667	68384	3393.66667	70750	3135.66667
66078.3333	3804.33333	63712.3333	4062.33333	68571.3333	3515.66667
60829.6667	4123.33333	70522	3206.33333	66767.6667	3657.33333
60231.6667	4143.66667	70187.6667	3252	66078.3333	3804.33333
66205.3333	3773.66667	65850.3333	3875	66433.3333	3703
68009	3632	65516	3920.66667		
70187.6667	3252	60231.6667	4143.66667		
68343.3333	3586.33333	66205.3333	3773.66667		
70522	3206.33333	63114.3333	4082.66667		
63712.3333	4062.33333	60829.6667	4123.33333		
63114.3333	4082.66667				
65516	3920.66667				
55149.6667	4209.33333				
65850.3333	3875				



**Figura 15.** Comparación gráfica de frentes para *o2p20\_Tri* (valores de GMI).

De acuerdo a lo observado en la Tabla 2, la principal diferencia entre ambos algoritmos radica en la cantidad de soluciones obtenidas. Por otro lado, en Figura 15, se aprecia que los frentes aproximados producidos por ambas variantes son relativamente cercanos al frente óptimo. De igual manera, es posible observar que el frente de la variante con SSD incluye soluciones no contempladas en la que usa CD.

## 6.5 Evaluación de desempeño

Con el propósito de evaluar el desempeño del algoritmo Tri-NSGA-II-SSD y el impacto de la incorporación de SSD como estimador de densidad, se realizó un experimento que consistió en la resolución un conjunto de 10 instancias de distintos tamaños usando los algoritmos Tri-NSGA-II-SSD y Tri-NSGA-II-CD.

Para medir el desempeño de los algoritmos se calcularon las siguientes métricas:

- Hipervolumen (*HyperVolume*, HV por sus siglas en inglés) [41]: es un indicador del volumen del espacio de  $n$  dimensiones dominado por las soluciones el frente de referencia. La magnitud de dicho espacio indica la cercanía al Frente de Pareto. Por ello, es deseable que se obtengan valores grandes.
- Dispersión generalizada (*Generalized Spread*, GS por sus siglas en inglés) [42]: es el promedio de las distancias que existen entre los puntos del frente de referencia y el vecino más cercano. Los valores pequeños son deseables, ya que significa que las soluciones del frente de referencia están bien distribuidas.

### 6.5.1 Generación de frentes de referencia

Cada uno de los algoritmos fue ejecutado un total de 30 veces por instancia. Con el propósito de contar con una aproximación al frente óptimo de cada instancia, se generó un frente de referencia para cada una de ellas. Cada frente de referencia está conformado por el conjunto de soluciones no dominadas de la combinación de los 30 frentes obtenidos luego de la ejecución de cada algoritmo.

Los frentes de referencia obtenidos fueron utilizados para el cálculo de las métricas de hipervolumen (*HyperVolume*) y dispersión generalizada (*Generalized Spread*). De manera que, luego de resolver 30 veces cada una de las instancias con ambos algoritmos, se calcularon 60 valores (30 de hipervolumen y 30 de dispersión generalizada) por algoritmo.

Luego de obtener los valores de ambas métricas por instancia, se calcularon los valores de la mediana ( $M_e$ ) y el rango intercuartílico (*IQR*) de dichas métricas para cada instancia. Debido a que el rango intercuartílico y la mediana son menos sensibles a valores extremos en comparación con la media y la desviación estándar, se optó por calcular dichas medidas de tendencia.

En la siguiente sección, se muestran los resultados obtenidos luego de realizar la evaluación de desempeño de los algoritmos.

## 6.5.2 Resultados de evaluación

Con el propósito de validar los resultados obtenidos, se les aplicó una prueba de Wilcoxon con un nivel de confianza de 95%. En las tablas que se muestran a continuación es posible percibir un cambio de tonalidad en ciertos resultados. Aquellos resultados cuyo resultado es significativo (de acuerdo a la prueba) se encuentran resaltados.

**Tabla 3.** Resultados de la evaluación de Tri-NSGA-II-CD y Tri-NSGA-II-SSD para la métrica de hipervolumen.

<i>HyperVolume</i>				
<b>Instancia</b>	<b>Tri-NSGA-II-CD</b>		<b>Tri-NSGA-II-SSD</b>	
	$M_e$	IQR	$M_e$	IQR
o2p20_Tri	0.238811441	0.044851738	0.163710911	0.069198901
o4p25_0Tri	0.264973213	0.05253274	0.20611812	0.096375359
o4p25_1Tri	0.209264293	0.079859039	0.146343058	0.100503483
o4p25_2Tri	0.259096812	0.066920563	0.209574394	0.121363497
o4p25_3Tri	0.172032511	0.03686516	0.161246173	0.083199129
o4p25_4Tri	0.26040734	0.097399762	0.213590777	0.122254636
o4p25_5Tri	0.227411752	0.064154667	0.180578485	0.087828932
o9p100_0Tri	0.072317856	0.078978551	0.059251828	0.066776339
o9p100_1Tri	0.081862723	0.079925586	0.051945505	0.041187613
o9p100_2Tri	0.124196526	0.092999414	0.084485714	0.0633435

**Tabla 4.** Resultados de la evaluación de Tri-NSGA-II-CD y Tri-NSGA-II-SSD para la métrica de dispersión generalizada.

<i>Generalized Spread</i>				
<b>Instancia</b>	<b>Tri-NSGA-II-CD</b>		<b>Tri-NSGA-II-SSD</b>	
	$M_e$	IQR	$M_e$	IQR
o2p20_Tri	0.95787788	0.45810173	0.71617801	0.1342707
o4p25_0Tri	0.41706173	0.06932411	0.34831437	0.06777055
o4p25_1Tri	0.403559	0.10486043	0.32747673	0.07611996
o4p25_2Tri	0.35834772	0.08948966	0.31741646	0.06983148
o4p25_3Tri	0.43841931	0.09865111	0.33124594	0.08006976
o4p25_4Tri	0.41285987	0.08332076	0.30090141	0.05669353
o4p25_5Tri	0.33274883	0.05214539	0.30352366	0.04080783
o9p100_0Tri	0.19171175	0.03487414	0.19084287	0.0275921
o9p100_1Tri	0.18635716	0.04270058	0.19014306	0.03099422
o9p100_2Tri	0.18869284	0.03584193	0.19065208	0.03091227

En la Tabla 3 puede observarse que Tri-NSGA-II-CD es mejor con diferencia significativa en la resolución de las 10 instancias de prueba para la métrica de hipervolumen.

Por otro lado, en la Tabla 4 es posible observar que Tri-NSGA-II-SSD tiene mejor desempeño en 7 de 10 instancias, en todas ellas con diferencia significativa.

Para validar los resultados obtenidos, se realizó una prueba de *Rankings* de Friedman con los valores obtenidos para ambas métricas. En este tipo de pruebas, se prefiere un valor menor de *Ranking*.

**Tabla 5.** Rankings de Friedman de Tri-NSGA-II-CD y Tri-NSGA-II-SSD para las métricas de hipervolumen y dispersión generalizada.

<b><i>HyperVolume</i> (p-value: 4.43753E<sup>-11</sup>)</b>		<b><i>Generalized Spread</i> (p-value: 6.84371E<sup>-11</sup>)</b>	
<b>Algoritmo</b>	<b>Ranking</b>	<b>Algoritmo</b>	<b>Ranking</b>
Tri-NSGA-II-CD	1.3066666666666682	Tri-NSGA-II-CD	1.7366666666666628
Tri-NSGA-II-SSD	1.693333333333331	Tri-NSGA-II-SSD	1.2633333333333354

Como puede observarse en la Tabla 5, los resultados obtenidos con ambas métricas fueron validados mediante la aplicación de la prueba de Friedman, en donde puede observarse que para la métrica de hipervolumen Tri-NSGA-II-CD tiene menor ranking y que para la métrica de dispersión generalizada, la variante con SSD tiene un ranking menor.

# Capítulo 7

## Conclusiones y trabajos futuros

Este trabajo de tesis cumple satisfactoriamente con los objetivos fijados en el momento de su concepción. En la presente sección se detallan las principales contribuciones del mismo, incluyendo modelos, mecanismos, productos científicos y posibles trabajos futuros.

### 7.1 Contribuciones

Las principales contribuciones de este trabajo de tesis se enuncian a continuación:

- Un algoritmo capaz de resolver instancias pequeñas del problema de selección de cartera de proyectos con parámetros difusos triangulares de forma exhaustiva (ver Sección 6.2).
- Un nuevo modelo matemático para la versión multiobjetivo del problema de selección de cartera de proyectos que incorpora números difusos triangulares como mecanismo de representación de incertidumbre presente en sus parámetros (ver Capítulo 4).
- Dos algoritmos basados en la metodología clásica del algoritmo NSGA-II que permiten la resolución del problema de selección de cartera de proyectos con parámetros difusos triangulares: Tri-NSGA-II-CD y Tri-NSGA-II-SSD (ver Capítulo 5).

## **7.2 Productos científicos**

Los productos científicos que se generaron durante el desarrollo de este trabajo se describen a continuación:

- Participación en el Congreso Mexicano de Inteligencia Artificial 2021 (COMIA2021), organizado por el Consejo Mexicano de Inteligencia Artificial con la exposición del artículo “Modelando y Optimizando el Problema de Selección de Portafolio de Múltiples Objetivos y Parámetros Difusos Triangulares”.
- Este trabajo de tesis para obtener el grado de Maestro en Ciencias de la Computación.

## **7.3 Trabajos a futuro**

Como posibles trabajos a futuro se contemplan:

- La implementación de variantes de estado estable de los algoritmos propuestos y la evaluación de su desempeño considerando calidad y eficiencia.
- Estudiar la incorporación de heurísticas o búsquedas locales.
- Estudiar la incorporación de nuevos operadores genéticos.

# Referencias

---

- [1] P. Ekel, W. Pedrycz y J. Pereira, *Multicriteria Decision-Making Under Conditions Of Uncertainty: A Fuzzy Set Perspective*, Wiley, 2020.
- [2] F. Balderas Jaramillo, *Modelando la imprecisión del problema de cartera de proyectos con filosofía gris*, Tijuana, Baja California, México: Tecnológico Nacional de México, 2018.
- [3] A. Estrada-Padilla y Otros, «Modeling and Optimizing the Multi-Objective Portfolio Optimization Problem with Trapezoidal Fuzzy Parameters,» *Mathematical and Computational Applications*, vol. 26, n° 2, p. 36, 2021.
- [4] S. French, «Uncertainty and imprecision: Modelling and analysis,» *Journal of the Operational Research Society*, vol. 46, n° 1, pp. 70-79, 1995.
- [5] G. J. Klir y B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall PTR, 1995.
- [6] G. Dantzig, «Linear programming under uncertainty,» *Management Science*, vol. 1, n° 2, pp. 197-207, 1955.
- [7] M. J. Kochenderfer, *Decision Making Under Uncertainty*, Cambridge, Massachusetts: MIT Press, 2015.
- [8] H. Zhang y D. Liu, *Fuzzy Modeling and Fuzzy Control*, Boston: Birkhäuser, 2006.
- [9] F. Pérez y T. Gómez, «Multiobjective Project Portfolio Selection with Fuzzy Constraints,» *Annals of Operations Research*, vol. 245, n° 1-2, pp. 7-29, 2016.
- [10] L. Yong-Jun y Z. Wei-Guo, «Fuzzy Portfolio Optimization Model under Real Constraints,» *Insurance: Mathematics and Economics*, vol. 53, n° 3, pp. 704-11, 2013.
- [11] F. Balderas, E. Fernández, C. Gómez-Santillán, N. Rangel-Valdez y L. Cruz, «An Interval-Based Approach for Evolutionary Multi-Objective Optimization of Project Portfolios,» *International Journal of Information Technology & Decision Making*, vol. 18, n° 4, pp. 1317-1358, 2019.
- [12] C. Yann y P. Siarry, *Multiobjective Optimization: Principles and Case Studies*, Springer, 2011.
- [13] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, 2001.
- [14] A. Ben-Tal, «Characterization of Pareto and Lexicographic Optimal Solutions,» *Multiple Criteria Decision Making Theory and Application*, vol. 177, pp. 1-11, 1980.

- [15] D. Simon, Evolutionary optimization algorithms: biologically-Inspired and population-based approaches to computer intelligence, John Wiley & Sons Inc, 2013.
- [16] H. Markowitz, «Portfolio Selection,» *The Journal of Finance*, vol. 7, n° 1, p. 77, 1952.
- [17] A. F. Carazo y otros, «Solving a Comprehensive Model for Multiobjective Project Portfolio Selection,» *Computers & Operations Research*, vol. 37, n° 4, pp. 630-39, 2010.
- [18] W. Pedrycz, P. Ekel y R. Parreiras, Fuzzy multicriteria decision-making: models, methods and applications, Wiley, 2011.
- [19] P. Ekel, «Approach to decision making in fuzzy environment,» *Computers and Mathematics with Applications*, vol. 37, n° 1, pp. 59-71, 1999.
- [20] I. N. Durbach y S. Theodor J., «Modeling Uncertainty in Multi-Criteria Decision Analysis,» *European Journal of Operational Research*, vol. 223, n° 1, pp. 1-14, 2012.
- [21] J. Vahidi y S. Rezvani, «Arithmetic Operations on Trapezoidal Fuzzy Numbers,» *Journal of Nonlinear Analysis and Application*, vol. 2013, pp. 1-8, 2013.
- [22] L. A. Zadeh, «Fuzzy Sets,» *Information and Control*, vol. 8, n° 3, pp. 338-53, 1965.
- [23] H. T. Nguyen y E. Walker, A first course in fuzzy logic, CRC Press, 2019.
- [24] S. H. Chen y C. H. Hsieh, «Representation , Ranking , Distance , and Similarity of L-R type fuzzy number and Application,» *Australian Journal of Intelligent*, vol. 6, n° 4, pp. 217-29, 2000.
- [25] S. Yao y otros, «A Multi-Objective Dynamic Scheduling Approach Using Multiple Attribute Decision Making in Semiconductor Manufacturing,» *International Journal of Production Economics*, vol. 130, n° 1, pp. 125-33, 2011.
- [26] R. Marti, Handbook of heuristics, Berlin: Springer Berlin Heidelberg, 2018.
- [27] J. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, Londres: MIT Press, 1992.
- [28] D. A. Coley, An introduction to genetic algorithms for scientists and engineers, World Scientific, 2010.
- [29] R. Larson y B. Farber, Elementary Statistics: Picturing the World, Prentice Hall, 2012.
- [30] A. Aczel y J. Sounderpandian, Complete Business Statistics, USA: McGraw-Hill Primis, 2008.

- [31] F. Wilcoxon, «Individual Comparisons by Ranking Methods,» *Biometrics Bulletin*, vol. 1, n° 6, p. 80, 1945.
- [32] R. H. Riffenburg, *Statistics in Medicine*, USA: Elsevier, 2005.
- [33] S. Bastiani Medina, *Solución de Problemas de Cartera de Proyectos Públicos a partir de información del Ranking de Prioridades (Tesis doctoral)*, Tijuana, Baja California, México: Tecnológico Nacional de México, 2017.
- [34] W. Gao y otros, «Artificial Bee Colony Algorithm Based on Novel Mechanism for Fuzzy Portfolio Selection,» *IEEE Transactions on Fuzzy Systems*, vol. 27, n° 5, pp. 966-78, 2019.
- [35] M. B. Kar y otros, «A New Bi-Objective Fuzzy Portfolio Selection Model and Its Solution through Evolutionary Algorithms,» *Soft Computing*, vol. 23, n° 12, pp. 4367-81, 2019.
- [36] C. Li y J. Jin, «Fuzzy Portfolio Optimization Model with Fuzzy Numbers,» *Nonlinear Mathematics for Uncertainty and its Applications*, vol. 100, pp. 557-65, 2011.
- [37] K. Deb y otros, «A fast and elitist multiobjective genetic algorithm: NSGA-II,» *IEEE Transactions on Evolutionary Computation*, vol. 6, n° 2, pp. 182-97, 2002.
- [38] A. Santiago y otros, «A Novel Multi-Objective Evolutionary Algorithm with Fuzzy Logic Based Adaptive Selection of Operators: FAME,» *Information Sciences*, vol. 471, pp. 233-51, 2019.
- [39] D. H. Ackley y T. M. Mitchell, *Connectionist Machine for Genetic Hillclimbing*, Springer US, 1987.
- [40] G. Syswerda, «Uniform Crossover in Genetic Algorithms,» de *Proceedings of the 3rd International Conference on Genetic Algorithms*, US, 1989.
- [41] L. While y otros, «A Fast Way of Calculating Exact Hypervolumes,» *IEEE Transactions on Evolutionary Computation*, vol. 16, n° 1, pp. 86-95, 2012.
- [42] A. Zhou y otros, «Combining Model-based and Genetics-based Offspring Generation for Multi-objective Optimization Using a Convergence Criterion,» de *IEEE International Conference on Evolutionary Computation*, 2006.
- [43] A. Medaglia y otros, «A multiobjective evolutionary approach for linearly constrained project selection under uncertainty,» *European Journal of Operational Research*, vol. 173, n° 3, pp. 869-894, 2007.