

Diversidad en los Algoritmos Evolutivos

Carlos Segura^{a,*}, Joel Chacón Castillo^a

^a*Area of Computer Science, Centre for Research in Mathematics (CIMAT), Callejón Jalisco s/n, Mineral de Valenciana, Guanajuato, Guanajuato 36240, Mexico*

Abstract

Differential Evolution is an efficient metaheuristic for continuous optimization that suffers from the curse of dimensionality. A large amount of experimentation has allowed researchers to find several potential weaknesses in Differential Evolution. Some of these weaknesses do not significantly affect its performance when dealing with low-dimensional problems, so the research community has not paid much attention to them. The aim of this paper is to provide a better insight into the reasons of the curse of dimensionality and to propose techniques to alleviate this problem. Two different weaknesses are revisited and schemes for dealing with them are devised. The schemes increase the diversity of trial vectors and improve on the exploration capabilities of Differential Evolution. **Some important mathematical properties induced by our proposals are studied and compared against the ones of related schemes. Experimentation with a set of problems with up to 1000 dimensions and with several variants of Differential Evolution shows that the analyzed weaknesses significantly affect the performance of Differential Evolution when used on high-dimensional optimization problems. The gains of the proposals appear when too exploitative schemes are used. Our proposals enable the achievement of high-quality solution with small populations, meaning that the most significant advantages emerge when dealing with large-scale optimization problems, where the benefits of using small populations have been previously shown.**

Keywords: Diversidad, Multi-objetivo, Convergencia Prematura, Evolución Diferencial

1. Introducción

Los Algoritmos Evolutivos (Evolutionary Algorithms EA) son considerados como uno de los enfoques con mayor eficacia para resolver distintas categorías de optimización. Particularmente, se han aplicado en problemas tanto de dominio continuo [21] como de dominio discreto. Especialmente, los EAs son aplicados para resolver problemas complejos cuyo enfoque determinístico es complicado o imposible [9]. Además, diversas variantes se han utilizado y aplicado en muchos campos, como es en ciencia, economía e ingeniería. Actualmente, los EAs son plenamente conocidos como metaheurísticas [21]. A pesar del éxito que tienen los

*Corresponding author. Tel.: + 52 473 732 7155

Email addresses: carlos.segura@cimat.mx (Carlos Segura), joel.chacon@cimat.mx (Joel Chacón Castillo)

EAs, existe una dificultad en su adaptación o configuración ante nuevos problemas. Una dificultad popular en el diseño de un EA es obtener un balanceo propio entre exploración y explotación [23]. Sin embargo, no siempre son comprendidas las implicaciones al mantener un grado de diversidad en este balanceo y como son promovidas las exploración y explotación [49]. Desde su inicio los EAs han presentado problemas de convergencia siendo como una desventaja importante [49]. La convergencia prematura es originada cuando todos los miembros de la población están ubicados en una parte reducida del espacio de búsqueda, esta región es distinta a la región óptima y además los componentes seleccionados no son suficientes para escapar de esta región. Basado en esto se han desarrollado varias estrategias para aliviar este problema. Además en través de varios estudios se ha revelado que mantener una población diversa en un requisito previo para evitar la convergencia prematura [49]. Sin embargo, si la población es muy diversa, entonces un grado adecuado de explotación podría ser prevenido, resultando en una convergencia lenta y por lo tanto soluciones de baja calidad. Por esta razón, Mahfoud [11] utilizó el concepto de diversidad útil, con cual se refiere a la cantidad de diversidad que resulta en soluciones de alta calidad. En la literatura existen distintas formas para aliviar el problema de convergencia prematura [39]. En los 90s, la mayoría de estrategias para aliviar la convergencia prematura se centraron en modificar el esquema de selección de padres. La principal razón es que en esa época la mayoría de esquemas eran generacionales, por lo tanto la presión de selección estaba definida principalmente en la selección de padres. Sin embargo, se descubrió que tratando de aliviar la convergencia prematura donde únicamente sea considerada la selección de padres no fue suficiente [5]. Posteriormente, la una gran cantidad de EAs incorporaron una fase de reemplazo que abandonaron al menos parcialmente a los métodos generacionales iniciales. Basado en esto muchos autores descubrieron la posibilidad de incorporar métodos para aliviar el problema de convergencia prematura [49]. Es importante considerar que aún cuando los métodos de reemplazamiento generacionales [12] fueron suficientemente populares, algunos autores ya habían tomado en cuenta esta estrategia [31]. Sin embargo, con la efectividad de elitismo y otras estrategias de reemplazo, el número de esquemas que adoptaron estos principios crecieron de forma considerable [30].

Un aspecto importante de la convergencia prematura es que depende completamente de la cantidad de tiempo y/o generaciones asignados a las ejecuciones del EA, es decir el criterio de paro. En su lugar, un EA debería ser ejecutado para resolver un problema dado por un tiempo definido y éste debería proporcionar soluciones prometedoras. A pesar de esto, es sorprendente que la mayoría de métodos que se han propuesto para aliviar la desventaja de la convergencia prematura no consideran el criterio de paro el cual es asignado por el usuario para alterar su comportamiento interno. Esto significa que, dependiendo en el criterio de paro, distintas parametrizaciones podrían ser requeridas. Como resultado, para criterio de paro distinto, el usuario debería estudiar el efecto de distintos parámetros. Un ejemplo popular de esto es *El Torneo de Selección Restringida (TSR)* [49], este método retrasa la convergencia de los EAs. Específicamente, este método incorpora un parámetro que puede ser utilizado para alterar el balanceo entre exploración e intensificación. Sin embargo, la pérdida de diversidad y posteriormente el balanceo entre exploración e intensificación no

dependen únicamente en este parámetro, por lo tanto distintos valores deberían ser utilizados para cada problema y además para cada criterio de paro. El principio básico de las técnicas para la preservación de la diversidad y que afectan a la fase de reemplazo se basa en que el efecto de diversificar a los sobrevivientes induce un mayor grado de exploración. Esto se debe a varios aspectos importantes, principalmente una población grande mantiene varias regiones del espacio de búsqueda. Además los operadores de cruce tienden a ser más explorativos cuando están involucrados individuos distantes [18].

Entre las distintas categorías de EAs, Evolución Diferencial (Differential Evolution - DE) es una de las estrategias más efectivas para lidiar con problemas de optimización continua [46]. De hecho, esta categoría de algoritmos han sido los ganadores en varias competencias de optimización [10]. Similarmente a otros EAs, DE está inspirado en el proceso de evolución natural, y además involucra la aplicación de mutación, recombinación y selección. La principal característica de DE es que éste considera las diferencias de los vectores que están presentes en la población con el motivo de explorar el espacio de búsqueda. En este sentido DE es similar a los optimizadores *Nelder-Mead* [37] y a la *Búsqueda Aleatoria Controlada (BAC)* [42]. A pesar de la efectividad de DE, existen varias debilidades que han sido detectadas y resueltas de forma parcial que por lo tanto han generado extensiones a la variante estándar de DE [10]. Algunos de los problemas más conocidos es la sensibilidad de la parametrización [53], la aparición de estancamiento debido a las capacidades de exploración reducidas [43, 29] y la convergencia prematura [52]. Desde la aparición de DE, varias críticas se hicieron debido a su falta de capacidad para mantener un grado de diversidad suficiente dado a la elevada presión de selección [43]. Por lo tanto, se han generado varias extensiones de DE para aliviar la convergencia prematura, como la adaptación de parámetros [52], auto-adaptación de la diversidad en la población [51] y estrategias de selección con una menor presión de selección [43]. Algunos de los últimos estudios en el diseño metaheurísticas poblacionales [49] mostraron que controlando explícitamente la diversidad es particularmente útil para obtener un propio balanceo entre el grado de exploración y de intensificación. Nuestra hipótesis es que introduciendo un mecanismo para la preservación de la diversidad resulta en un balanceo adecuado entre exploración e intensificación, que a su vez proporciona soluciones de alta calidad en ejecuciones a largo plazo.

Principalmente en este capítulo se explican dos propuestas, primeramente DE con Mantenimiento de Diversidad Mejorado (DE with Enhanced Diversity Maintenance-DE-EDM), el cual integra un principio similar en DE y el operador Dinámico basado en el Cruce Binario Simulado (Dynamic Simulated Binary Crossover-DSBX).

Our novel proposal, which is called DE with Enhanced Diversity Maintenance (DE-EDM), integrates a similar principle into DE. El resto de este capítulo está organizado de la siguiente forma.

2. Mecanismos para evitar la convergencia prematura

3. Diseño de Evolución Diferencial basado en diversidad

3.1. Evolución Diferencia: Conceptos Básicos

Esta sección esta dedicada para repasar la variante clásica de DE y para introducir algunos de los mas importantes términos utilizados en el campo de DE. El clásico esquema DE es identificado como DE/rand/1/bin el cual ha sido extensamente utilizado para generar más variantes complejas [10]. De hecho, nuestra propuesta también extiende a la clásica versión DE. Originalmente DE fue propuesta como un método de búsqueda directo para optimización continua mono-objetivo. El conjunto de variables involucradas en el planteamiento de un problema son dados como un vector de la forma $\vec{X} = [x_1, x_2, \dots, x_D]$, donde D es la dimensión del problema. En optimización continua, cada x_i es un número real, además son proporcionadas restricciones de caja, es decir, existe un límite inferior (a_i) y un límite superior (b_i) para cada variable. El objetivo de un proceso de optimización es obtener un vector \vec{X}^* el cual minimiza una función objetivo dada, esto matemáticamente es definido por $f : \Omega \subseteq \mathbb{R}^D \rightarrow \mathbb{R}$. En el caso de la restricción de caja $\Omega = \prod_{j=1}^D [a_j, b_j]$.

DE es un algoritmo estocástico basado en una población, por lo tanto éste involucra iterativamente a un conjunto de soluciones candidatas. En DE dichas soluciones candidatas son usualmente conocidas como vectores. En la variante básica de DE, para cada miembro de la población conocidos como *vectores objetivo* es generado un nuevo vector conocido como *vector mutado*. Entonces, el vector mutado es combinado con el vector objetivo para generar al *vector de prueba*. Finalmente, una fase de selección es aplicada para seleccionar a los vectores sobrevivientes. De esta forma, transcurren las generaciones hasta cumplir el criterio de paro. El i -ésimo vector de la población en la generación G es definido como $\vec{X}_{i,G} = [x_{1,i,G}, x_{2,i,G}, \dots, x_{D,i,G}]$. A continuación se explica en más detalle cada componente de DE.

3.1.1. Inicialización

DE usually starts the optimization process with a randomly initiated population of NP vectors. Since there is commonly no information about the performance of different regions, uniform random generators are usually applied. Hence, the j th component of the i th vector is initialized as $x_{j,i,0} = a_j + rand_{i,j}[0, 1](b_j - a_j)$, where $rand_{i,j}[0, 1]$ is an uniformly distributed random number lying between 0 and 1.

3.1.2. Mutación

Para cada vector objetivo un vector mutado es creado, varias estrategias para realizar este procedimiento han sido propuestas. En la variante clásica de DE se aplica la estrategia rand/1. En este caso, es creado el vector mutado $V_{i,G}$ de la siguiente forma:

$$\vec{V}_{i,G} = \vec{X}_{r1,G} + F \times (\vec{X}_{r2,G} - \vec{X}_{r3,G}) \quad r1 \neq r2 \neq r3 \quad (1)$$

Los índices $r1, r2, r3 \in [1, NP]$ distintos enteros seleccionados de forma aleatoria en el rango $[1, NP]$. Además, estos índices son distintos al índice i . Es importante tomar en cuenta que la diferencia entre los vectores es escalada por medio del parámetro F , el cual se define usualmente en el intervalo $[0, 4, 1]$. La diferencia escalada es agregada al tercer vector, por lo tanto los vectores mutados son similares a los vectores objetivo cuando el grado de diversidad es poco y las diferencias son pequeñas. Como resultado, es importante mantener un grado de diversidad mínimo en DE.

3.1.3. Cruza

En orden con el objetivo de combinar la información de distintas soluciones candidatas y con el propósito de incrementar la diversidad es aplicado el operador de cruce. Específicamente, cada vector objetivo $\vec{X}_{i,G}$ es mezclado con su correspondiente vector mutado $V_{i,G}$ para generar un vector de prueba $\vec{U}_{i,G} = [u_{1,i,G}, u_{2,i,G}, \dots, u_{D,i,G}]$. La estrategia de cruce mas típica es conocida como cruce *binomial*, el cual funciona de la siguiente forma:

$$\vec{U}_{j,i,G} = \begin{cases} \vec{V}_{j,i,G}, & \text{si } (rand_{i,j}[0, 1] \leq CR \text{ or } j = j_{rand}) \\ \vec{X}_{j,i,G}, & \text{de otra forma} \end{cases} \quad (2)$$

donde $rand_{i,j}[0, 1]$ es un número uniformemente distribuido, j_{rand} es un índice aleatoriamente seleccionado el cual asegura que $\vec{U}_{i,G}$ genera al menos un componente de $\vec{V}_{i,G}$ y $CR \in [0, 1]$ es la razón de cruce.

3.1.4. Selección

Finalmente, se aplica una selección glotona para determinar a los sobrevivientes de la siguiente generación. Cada vector de prueba es comparado con su correspondiente vector objetivo y el mejor es el que sobrevive:

$$\vec{X}_{j,i,G+1} = \begin{cases} \vec{U}_{i,G}, & \text{si } f(\vec{U}_{i,G}) \leq f(\vec{X}_{i,G}) \\ \vec{X}_{i,G}, & \text{de otra forma} \end{cases} \quad (3)$$

Hence, each population member either gets better or remains with the same objective value in each generation. Since members never deteriorate, it is considered to be a selection with high pressure. Note that in case of a tie, the trial vector survives.

3.2. Diversidad en Evolución Diferencial

Los algoritmos basados en DE son altamente susceptibles a la pérdida de diversidad debido a la estrategia de selección agresiva. Sin embargo, se han desarrollado varios análisis para lidiar con este problema. Desde que se conocen las implicaciones de cada parámetro en la diversidad, una alternativa es la estimación teórica de los valores adecuados en DE [52]. Alternativamente, se han desarrollado algunos análisis donde es considerado el efecto de los vectores de diferencia en la mutación [33]. Estos análisis y otros estudios empíricos

basados en la cruza permitieron concluir que ciertos tipos de movimientos deberían ser deshabilitados para retrasar la convergencia [35]. En este último estudio varía el tipo de movimientos aceptados a lo largo de la ejecución. Específicamente, esto descarta movimientos menores a un umbral el cual es decrementado conforme transcurren las generaciones. Se han propuesto otras formas de alterar el procedimiento en que se aceptan los movimientos [6]. Es importante notar que este tipo de métodos tienen similitudes con nuestra propuesta en el sentido de que las decisiones están basadas por el número de generaciones transcurridas. Sin embargo, nuestro método opera en la estrategia de reemplazo y no en la fase de mutación. Mas aún, estos métodos no consideran de forma explícita las diferencias que aparecen en la población entera. En su lugar, las restricciones son aplicadas a las diferencias que aparecen en la fase de reemplazo.

Una alternativa distinta reside en alterar el operador de selección [43]. Particularmente, se relaja la presión de selección a través de una selección probabilística con el propósito de mantener la diversidad en la población y consecuentemente permitir escapar de la base de atracción de un óptimo local. Sin embargo este método es muy sensible a transformaciones desde que esta estrategia considera la aptitud para definir las probabilidades para aceptar un individuo mutado. En este caso las decisiones no se basan en las generaciones transcurridas.

Finalmente, en el algoritmo *Diversidad de la Población Auto-Mejorado* (*Auto-Enhanced Population Diversity* - AEPD), la diversidad es explícitamente medida y esto es un disparo a un mecanismo para diversificar a la población cuando se detecta poca diversidad en la población [51]. También ya se han propuesto estrategias con principios similares pero con esquemas de perturbación distintos.

Es importante notar que las mejores variantes-DE de las competencias no utilizan estas modificaciones y que la mayoría de estas extensiones no han sido implementadas en los herramientas de optimización más utilizadas. Como resultado, estas extensiones no son ampliamente utilizadas por la comunidad a pesar de sus beneficios en ciertos casos.

3.3. Propuesta

Nuestra propuesta está motivada por dos trabajos significativos de ésta área cuyo propósito es el control de la diversidad en los EAs. Por su parte el primero es un estudio empírico desarrollado por Montgomery y otros [35], este trabajo presenta varios análisis empíricos los cuales confirman los problemas relacionados con la convergencia prematura.

Por otro lado, el segundo trabajo propuesto por Segura [44] y otros, proporciona mejoras significativas en el campo optimización combinatoria, en esta propuesta se desarrolla una novel estrategia de reemplazo nombrada *Reemplazo con Control de Diversidad Dinámico Basado en Varios Objetivos* (Replacement with Multi-objective based Dynamic Diversity Control - RMDDC) donde se controla el grado de diversidad con el criterio de paro y las generaciones transcurridas. Se obtuvieron beneficios por los métodos que incluyeron el RMDDC, por lo tanto y basados en las conclusiones de los trabajos previos, la propuesta de esta sección es

Algorithm 1 Esquema general del DE-EDM

```
1: Inicializar de forma aleatoria a la población con  $NP$  individuos, donde cada uno es distribuido de forma uniforme.
2:  $G = 0$ 
3: while El criterio de paro no sea alcanzado do
4:   for  $i = 1$  to  $NP$  do
5:     Mutación: Generar al vector mutado ( $V_{i,G}$ ) de acuerdo a la ecuación (1).
6:     Cruza: Utilizar la recombinación para generar al vector de prueba ( $U_{i,G}$ ) de acuerdo a la ecuación (2).
7:     Selección: Actualizar al vector elite ( $E_{i,G}$  en lugar de  $X_{i,G}$ ) de acuerdo a la ecuación (3).
8:     Reemplazo: Seleccionar a los vectores objetivo ( $X_{G+1}$ ) de acuerdo a la ecuación (2).
9:    $G = G + 1$ 
```

una novel variante de DE que incluye un mecanismo explícito el cual sigue uno de los principios del RMDDC. Este novel optimizador es nombrado *Evolución Diferencial con Mantinimiento Mejorado de Diversidad* (Differential Evolution with Enhanced Diversity Maintenance - DE-EDM) y su código fuente está disponible de forma gratuita ¹

La esencia de DE-EDM (ver algoritmo 1) es suficiente similar a la versión estándar de DE. De hecho la forma en que se crean los vectores de prueba no es modificado de forma significativa (líneas 5 y 6). La novedad de la propuesta es que incorpora una población elite (E) y una novel estrategia de reemplazo basada en la diversidad. En orden, para seleccionar a los miembros de la población, se aplica el reemplazo agresivo (glotón) de la versión original de DE (línea 7). Por otra parte, se considera otra estrategia de reemplazo (línea 8), la cual realiza la selección de los miembros que participarán en el siguiente procedimiento de selección, esto se realiza siguiendo el mismo principio que el RMDDC, es decir los individuos que contribuyen muy poco a la diversidad no deberán ser aceptados como miembros en la siguiente generación. En este sentido, no se utiliza la misma estrategia de selección agresiva que pertenece a DE para mantener a la población padre (X). En este orden para establecer una contribución aceptable de diversidad mínima para realizar la selección, son tomados en cuenta el criterio de paro y las generaciones transcurridas. Una de las principales debilidades del RMDDC es que su convergencia se retrasa de forma significativa. Por lo tanto, se realizan dos modificaciones al RMDDC para promover una convergencia acelerada. Primero, no son considerados los conceptos multi-objetivo, en su lugar se considera una selección mas agresiva. Segundo, también es considerada la población elite en la estrategia de reemplazo.

Nuestra estrategia de reemplazo (ver Algoritmo 2) funciona de la siguiente forma. Éste recibe a la población padre (vectores objetivo), la población de hijos (vectores de prueba) y a los vectores elite. En cada generación son seleccionados NP vectores para la siguiente población de padres. Primero, en base al número de evaluaciones a función (línea 2) es calculada una distancia mínima D_t deseada para mantener la diversidad. Entonces, son juntadas las tres poblaciones en un conjunto de miembros candidatos (línea 3). El conjunto de miembros candidatos continen vectores que podrían ser seleccionados para sobrevivir.

¹El código en C++ puede ser descargado en la siguiente dirección https://github.com/joelchaconcastillo/Diversity_DE_Research.git

Algorithm 2 Fase de Reemplazo

```
1: Entrada: Población (Vectores Objetivo), Hijos (Vectores de prueba), y Elite
2: Actualizar  $D_t = D_I - D_I * (n_{fes} / (0.95 * max\_n_{fes}))$ 
3:  $Candidatos = Población \cup Hijos \cup Elite$ .
4:  $Sobrevivientes = Penalizados \emptyset$ .
5: while  $Sobrevivientes < NP$  y  $|Candidatos| > 0$  do
6:    $Seleccionados =$  Seleccionar al mejor individuo de  $Candidatos$ .
7:   Eliminar  $Seleccionado$  de  $Candidatos$ .
8:   Copiar  $Seleccionado$  a  $Sobrevivientes$ .
9:   Encontrar los individuos de  $Candidatos$  cuya distancia a  $Seleccionados$  sea menor que  $D_t$  y moverlos al  $Penalizados$ . En esta parte se considera la distancia normalizada (Ecuación 4).
10: while  $|Sobrevivientes| < NP$  do
11:    $Seleccionado =$  Seleccionar al individuo de  $Penalizados$  con la mayor distancia al individuo mas cercano a  $Sobrevivientes$ .
12:   Eliminar  $Seleccionado$  de  $Penalizados$ .
13:   Copiar  $Seleccionado$  a  $Sobrevivientes$ 
14: return  $Survivors$ 
```

Entonces, el conjunto de individuos sobrevivientes y penalizados son inicializados por el conjunto vacío (línea 4). En orden para seleccionar a los NP sobreviviente (población padre de la siguiente generación) se repite un proceso iterativo (líneas 5 - 13). En cada paso es seleccionado el mejor individuo para sobrevivir del *Conjunto de Candidatos*, es decir al individuo que tiene la mejor aptitud, entonces es movido al *Conjunto de Sobrevivientes*. Entonces, los individuos que pertenecen al *Conjunto de Candidatos* cuya métrica de mínima distancia sea menor que D_t son transferidos *Conjunto de Penalizados* (línea 9).

La forma para calcular la distancia entre dos individuos es en base a la distancia Euclideana normalizada descrita en la ecuación 4, donde D es la dimensión del problema, y a_d, b_d son los límites menores y mayores de cada dimensión (d). En los casos donde *El conjunto de Candidatos* está vacío previamente a la selección de los NP individuos, el *Conjunto de Sobrevivientes* se llena seleccionando en cada iteración al individuo *Penalizado* con la mayor distancia al individuo más cercano al *Conjunto de Sobrevivientes* (líneas 10 - 13).

$$distance(x_i, x_j) = \frac{\sqrt{\sum_{d=1}^D \left(\frac{x_i^d - x_j^d}{b_d - a_d} \right)^2}}{\sqrt{D}} \quad (4)$$

En este orden con el propósito de completar la descripción es importante especificar la forma en que se calcula D_t y el para actualizar a los individuos elite. El resto del algoritmo es mantenido de igual forma que la clásica variante de DE. El valor de D_t es utilizado para alterar el grado entre exploración y explotación, por lo tanto éste parámetro debería depender en la etapa de optimización. Específicamente, este valor debería ser reducido conforme se alcanza el criterio de paro con el objetivo de promover un grado de intensificación. En nuestro esquema se requiere asignar un valor inicial para D_t (D_I). Así, similarmente que en [44], se calcula una reducción lineal de D_t considerando las evoluciones a función y el criterio de paro. Particularmente, en este trabajo, el criterio de paro es asigna en base a las evaluaciones a función. La reducción es calculada de tal forma que en el 95 % del máximo número de evaluaciones a función el valor de D_t es 0. Por lo tanto, la

diversidad no es considerada del todo en el restante 5%. Entonces, si max_nfe s es el máximo número de evaluaciones y nfe s es el número de evaluaciones transcurridas nfe s, entonces D_t puede ser calculado de la siguiente forma $D_t = D_I - D_I * (nfe / (0,95 * max_nfe))$.

La distancia inicial (D_I) afecta de forma considerable al rendimiento del DE-EDM. Si este parámetro es elevado, entonces el algoritmo tiene como objetivo maximizar la diversidad de la población en las primeras etapas de optimización, por lo tanto se genera una exploración adecuada la cual es muy importante, particularmente en varios tipos de problemas tales como altamente multi-modales y deceptivos. Entonces, se podría aliviar el efecto de la convergencia prematura. Un valor muy elevado de D_I podría inducir exploración excesivamente y por lo tanto una fase de intensificación podría no ser efectuada. Por otra parte, un valor muy pequeño de D_I podría evitar la fase de exploración, por lo tanto será más difícil evitar óptimos locales. El óptimo D_I podría variar dependiendo en el tipo problema y el criterio de paro. En su lugar, los problemas deceptivos y altamente multi-modales usualmente requieren valores más elevados que en los problemas unimodales. Sin embargo, en nuestra propuesta no se adapta un valor D_I para cada problema, por lo tanto con el propósito de analizar la estabilidad de este parámetro se realiza un análisis con diferentes valores de D_I en la sección de validación experimental.

Al igual que a la versión estándar DE, en nuestra propuesta DE-EDM se debe asignar una probabilidad de cruce (CR) y un factor de mutación (F).

El primero es quizás es más importante de acuerdo a varios estudios desarrollados por Montgomery y otros [34]. Estos autores probaron de forma empírica que valores extremos de CR resultan en un comportamiento muy distinto. Ellos explicaron que bajos valores de CR resultan en una búsqueda que es alineada con un pequeño número de ejes y además induce pequeños desplazamientos. Esto provoca una convergencia lenta y gradual que en algunos escenarios podría provocar un comportamiento robusto. Adicionalmente, valores elevados de CR podrían generar soluciones de mayor calidad con una menor probabilidad. Sin embargo, estas transformaciones provocan largos desplazamientos que podrían mejorar de forma significativa. De acuerdo a esto, en nuestra propuesta se emplean los dos principios, es decir, valores elevados y pequeños de CR como es mostrado en la ecuación 5.

$$CR = \begin{cases} Normal(0,2,0,1), & \text{si } rand[0,1] \leq 0,5 \\ Normal(0,9,0,1), & \text{de otra forma} \end{cases} \quad (5)$$

Siguiendo los principios de distintas variantes del SHADE [2, 7], se consideran las evaluaciones a función en el proceso de generación aleatorio del factor de mutación F . Particularmente, cada valor F es una muestra de una distribución Cauchy (Ecuación6).

$$Cauchy(0,5,0,5 * nfe / max_nfe) \quad (6)$$

Por lo tanto, en las primeras etapas de optimización los valores de F son generados de forma cercana a 0,5.

Conforme la ejecución transcurre, la función de densidad sufre una transformación gradual donde la varianza se incrementa, esto implica que son generados valores fuera del intervalo $[0,0,1,0]$ con una probabilidad alta. En los casos cuando los valores son mayores a 1,0, es utilizado un valor de 1,0. Si se genera un valor negativo, entonces este valor se vuelve a generar. Uno de los efectos de este enfoque es incrementar la probabilidad de generar valores elevados de F conforme transcurren las generaciones con el objetivo de evitar una convergencia en las últimas etapas de optimización.

3.4. Resultados

En esta sección se presenta la validación experimental. Especialmente, demostramos que los resultados de los algoritmos que pertenecen al estado-del-arte pueden ser mejorados controlando explícitamente la diversidad en el clásico DE. Particularmente, se consideraron los conjuntos de prueba del CEC 2016 y CEC 2017. Cada uno está compuesto de treinta distintos problemas. El estado-del-arte está compuesto por los algoritmos que alcanzaron los primeros lugares en cada año. Adicionalmente, se incluyó la versión estándar DE. Por lo tanto, los algoritmos considerados del CEC 2016 son el UMOEAs-II [19] y L-SHADE-EpSin [2], los cuales alcanzaron el primero y el segundo lugar respectivamente. Similarmente, los mejores algoritmos del CEC 2017 son el EBOwithCMAR [28] y el jSO [8].

Es importante destacar que el EBOwithCMAR es considerado como una mejora del UMOEAs-II. Adicionalmente, el jSO y el L-SHADE-EpSin pertenecen a la familia de SHADE. Todos estos algoritmos fueron probados con los dos conjuntos de prueba como es sugerido en [32]. Debido a que todos los algoritmos son estocásticos se realizaron 51 ejecuciones con distintas semillas.

En cada caso, el criterio de paso fue asignado a 25,000,000 evaluaciones a función. La evaluación de los algoritmos se realizó siguiendo los lineamientos de las competencias del CEC. Entonces, se asignó un error de 0 si la diferencia entre la mejor solución encontrada y la solución óptima era menor que 10^{-8} . Para cada algoritmo se utilizó la parametrización indicada por los respectivos autores, que son definidos a continuación:

- **EBOwithCMAR:** Para la parte EBO, el tamaño máximo de la población de $S_1 = 18D$, el tamaño mínimo de la población de $S_1 = 4$, el tamaño máximo de la población de $S_2 = 146,8D$, el tamaño mínimo de la población de $S_2 = 10$, el tamaño de la memoria histórica $H=6$. Para la parte de CMAR el tamaño de la población $S_3 = 4 + 3\log(D)$, $\sigma = 0,3$, $CS = 50$, la probabilidad de búsqueda local $pl = 0,1$ y $cfel_s = 0,4 * FE_{max}$.
- **UMOEAs-II:** Para la parte de MODE, el tamaño máximo de la población de $S_1 = 18D$, el tamaño mínimo de la población de $S_1 = 4$, el tamaño de la memoria histórica $H=6$. Para la parte del CMA-ES el tamaño de la población $S_2 = 4 + \lfloor 3\log(D) \rfloor$, $\mu = \frac{PS}{2}$, $\sigma = 0,3$, $CS = 50$. Para la búsqueda local, $cfel_s = 0,2 * FE_{max}$.

- **jSO**: El tamaño máximo de la población = $25\log(D)\sqrt{D}$, el tamaño de la memoria histórica $H = 5$, valor de mutación inicial de la memoria $M_F = 0,5$, probabilidad inicial de la memoria $M_{CR} = 0,8$, tamaño mínimo de la población = 4, valor inicial p-best = $0,25 * N$, valor final p-best = 2.
- **L-SHADE-EpSin**: Tamaño máximo de la población = $25\log(D)\sqrt{D}$, tamaño de la memoria histórica $H = 5$, valor de la mutación inicial de la memoria $M_F = 0,5$, probabilidad inicial de la memoria $M_{CR} = 0,5$, frecuencia inicial de la memoria $\mu_F = 0,5$, tamaño mínimo de la población = 4, valor inicial p-best = $0,25 * N$, valor final p-best = 2, generaciones de la búsqueda local $G_{LS} = 250$.
- **DE-EDM**: $D_I = 0,3$, tamaño de la población = 250.
- **Standard-DE**: tamaño de la población = 250 (mismos operadores que en DE-EDM).

Nuestro análisis experimental se desarrolló en base a la diferencia entre la solución óptima y la mejor solución obtenida. En orden para comparar los resultados estadísticamente, se siguió un procedimiento similar que el propuesto en [17].

Concretamente, en primer lugar se utilizó el test Shapiro-Wilk para comprobar si los resultados se ajustaban a una distribución Gaussiana. En los casos en que sí se ajustaban, se utilizó el test de Levene para comprobar la homogeneidad de las varianzas, procediendo con el test de ANOVA en caso positivo o con el de Welch en caso negativo. Por otro lado, para los casos que no se ajustaban a distribuciones Gaussianas, se utilizó el test de Kruskal-Wallis. En todos los casos se fijó el nivel de confianza al 95 %. Se considera que un algoritmo X es superior a un algoritmo Y , si el procedimiento anterior reporta diferencias significativas y si la media y mediana del hipervolumen obtenido por el método X son superiores a las obtenidas por el método Y . En las tablas 1 y 2 se presenta un resumen de los resultados obtenidos para el CEC 2016 y el CEC 2017 respectivamente. La columna etiquetada con “Símple Resuelto” muestra el número de funciones en que se obtuvo un error de cero en las 51 ejecuciones. Adicionalmente, la columna etiquetada con “Al menos una vez resuelto” muestra el número de soluciones que se resolvieron en al menos una ejecución. Practicamente nuestra propuesta resolvió al menos una vez todas las funciones (28 funciones) que pertenecen al conjunto de problemas del CEC 2017. Adicionalmente, fueron resueltas al menos una vez 21 funciones que pertenecen al CEC 2016. Esta es una diferencia sustancial con los resultados obtenidos por los algoritmos que pertenecen al estado-del-arte. Estos algoritmos obtuvieron los valores óptimos significativamente en menos funciones. En orden para confirmar la superioridad del DE-EDM, se implementaron las pruebas estadísticas por pares. La columna etiquetada con el símbolo \uparrow muestra que el número de veces en que cada método fue superior, mientras que la columna etiquetada con \downarrow cuenta el número de casos donde el método fue inferior. Finalmente, la columna etiquetada con \longleftrightarrow muestra el número de comparaciones cuyas diferencias no fueron significativas. Las pruebas estadísticas indican que el DE-EDM alcanzó los mejores resultados en los dos años. De hecho el número de veces en que nuestra propuesta ganó en el CEC 2016 y el CEC 2017 fue

de 77 y 88 respectivamente. Además el número de veces en que perdió fueron de 25 y 6 respectivamente. Adicionalmente, el último lugar alcanzado en los dos años fue por el algoritmo L-SHADE-Epsilon con 20 comparaciones positivas en el 2016 y 7 comparaciones positivas en el 2017. La última columna etiquetada con “Puntaje” muestra un análisis que fue propuesto en las competencias del CEC. Particularmente, este método de evaluación combina dos puntajes como se indica en la ecuación (7). Por lo tanto el puntaje final está compuesto por la suma $Score = Score_1 + Score_2$.

$$\begin{aligned} Score_1 &= \left(1 - \frac{SE - SE_{min}}{SE}\right) \times 50, \\ Score_2 &= \left(1 - \frac{SR - SR_{min}}{SR}\right) \times 50, \end{aligned} \quad (7)$$

donde, SE_{min} es la suma mínima de errores entre todos los algoritmos, y SE es la suma de errores dado un algoritmo $SE = \sum_{i=1}^{30} error_f_i$. Similarmente, SR_{min} es la suma mínima de los rangos entre todos los algoritmos, específicamente es la suma de cada rango en cada función para los algoritmos considerados $SE = \sum_{i=1}^{30} error_f_i$. Principalmente, nuestra propuesta alcanzó los mejores puntajes (100,00) en los dos años, demostrando su superioridad. Adicionalmente, la versión estándar de DE alcanzó resultados suficientemente buenos, de hecho obtuvo el tercer y el segundo lugar en los años 2016 y 2017 respectivamente. Esto muestra que el rendimiento de los algoritmos en el estado-del-arte es distinto al considerar ejecuciones a largo plazo. El algoritmo L-SHADE-Epsilon obtuvo un puntaje competitivo a pesar de que en el CEC del 2017 alcanzó el menor número de comparaciones positivas dentro de las pruebas estadísticas. Esto podría ocurrir desde que los puntajes estadísticos consideran la media y mediana de los errores. Es más el puntaje considera el rango y la media basado en el error.

Dado que nuestra propuesta está basada en el control explícito de la diversidad y con el objetivo de entender mejor su comportamiento, en la figura 1 se muestra la evolución de la diversidad a través de las evaluaciones a función. Particularmente, se ejecutó el DE-EDM con las funciones F_1 y f_{30} . Basado en sus propiedades la primera función se resuelve de forma sencilla (unimodal) y la segunda función es considerada como una de las más difíciles (híbrida). En la parte izquierda se muestra la diversidad que se mantiene en la población Elite. A pesar de que existen mecanismos para evitar la pérdida de diversidad en la población Elite, se puede observar que se mantiene un grado de diversidad de forma implícito en las dos funciones. Similarmente, la parte derecha corresponde a la diversidad de los vectores de prueba. Esto demuestra que se demuestra un grado de diversidad de forma explícito, es decir hasta el 95 % del total de evaluaciones a función.

En orden, con el fin de proporcionar resultados comparables, en las tablas 3 y 4 se reporta el mejor, peor, mediana, media, desviación estándar y razón de éxito. Particularmente, en estas tablas se observa que nuestra propuesta resuelve a todos los problemas unimodales. Además, varias funciones multimodales son aproximadas de forma aceptable. Principalmente, nuestra propuesta resolvió y mejoró significativamente

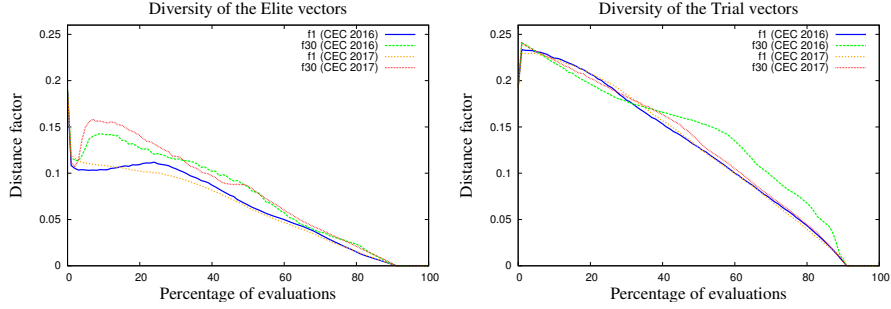


Figura 1: Promedio del DCN de las 51 ejecuciones con los problemas f_1 y f_{30} (CEC 2016 y CEC 2017). El factor de distancia inicial corresponde a $D_I = 0,3$.

Cuadro 1: Resumen de los resultados - CEC 2016

Algorithm	Siempre Resuelto	Resuelto al menos una vez	Pruebas Estadísticas			Puntaje
			↑	↓	↔	
EBOwithCMAR	8	14	35	56	59	50.28
jSO	9	17	47	51	52	55.43
UMOEAs-II	9	14	51	31	68	62.45
L-SHADE-Epsilon	7	13	20	71	59	50.12
DE-EDM	13	21	77	25	48	100.00
Standard-DE	11	19	50	46	54	56.29

varias funciones complejas (por ejemplo funciones computestas), que por otra parte no fueron resueltas por los algoritmos del estado-del-arte.

3.5. Análisis Empírico del factor de distancia inicial

En nuestra propuesta la diversidad es explícitamente promovida a través de varias etapas, y son promovidas por medio del factor de distancia inicial D_I . Por lo tanto, se analiza en detalle el efecto de este parámetro. Particularmente, se considera la configuración general de la validación experimental. Entonces, se consideraron varios factores de distancia inicial ($D_I = \{0,0, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1\}$).

En la figura 2 se muestra la razón de éxito promedio vs. el factor de distancia inicial (D_I). Principalmente se puede observar lo siguiente:

- Si la diversidad no es promovida ($D_I = 0,0$) entonces el rendimiento del algoritmo está comprometido.
- En este escenario la configuración ideal es de $D_I = 0,3$, a pesar de que aún existen soluciones de calidad en el rango $[0,1, 0,4]$.
- Si se incrementa la diversidad inicial entonces se observa un deterioro en la calidad de las soluciones.

Cuadro 2: Resumen de los resultados - CEC 2017

Algorithm	Siempre Resuelto	Resuelto al menos una vez	Pruebas Estadísticas			Puntaje
			↑	↓	↔	
EBOwithCMAR	9	18	34	46	70	37.14
jSO	8	15	29	55	66	29.30
UMOEAs-II	11	15	43	40	67	26.89
L-SHADE-Epsilon	8	19	7	81	62	32.78
DE-EDM	21	28	88	6	56	100.00
Standard-DE	12	21	56	29	65	42.91

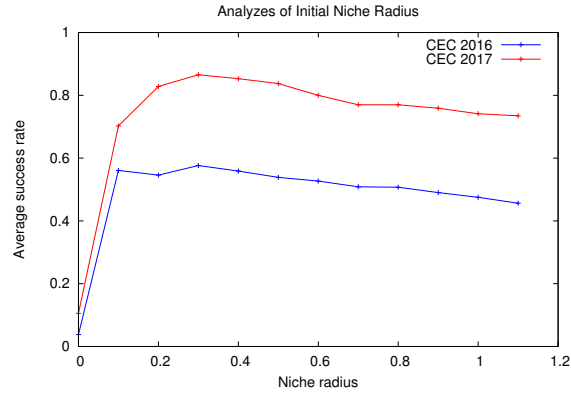


Figura 2: Razón de éxito promedio con distintos factores de distancia inicial con los problemas de prueba del CEC 2016 y CEC 2017, específicamente se considera una población de 250 individuos y 25,000,000 evaluaciones a función.

Finalmente, es importante aclarar que en base a varios estudios la calidad de las soluciones es afectado en un menor grado por el tamaño de la población que con el parámetro D_I .

4. Diseño de operadores de cruce basados en diversidad

Esta sección está destinada algunos de los trabajos mas importantes relacionados al ámbito de operadores de cruce y al concepto multi-objetivo. Primero, son definidos los paradigmas multi-objetivo mas importantes. Después de eso, se introducen algunas clasificaciones relevantes de los operadores de cruce. Finalmente, es discutido el operador de cruce SBX, el cual es utilizado de forma extensa en esta sección.

4.1. Algoritmos Evolutivos Multi-objetivo

En los últimos años, se han ideado una gran cantidad de MOEAs los cuales siguen distintos principios. En orden para mejor tener una mejor clasificación, se han propuesto varias taxonomías [45]. En base a sus principios de diseño, los MOEAs pueden ser basados en la dominancia de Pareto, indicadores y/o descomposición [40]. Todos estos algoritmos son suficientemente competitivos, por lo tanto en esta sección se

Cuadro 3: Resultados del DE-EDM con los problemas del CEC 2016

	Mejor	Peor	Mediana	Media	Sd	Razón de éxito
f_1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_5	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_6	0.00E+00	3.60E-02	4.00E-03	7.39E-03	1.15E-02	3.92E-01
f_7	2.00E-02	1.02E-01	5.90E-02	5.77E-02	4.93E-02	0.00E+00
f_8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_9	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{10}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{11}	0.00E+00	6.00E-02	0.00E+00	5.88E-03	1.90E-02	9.02E-01
f_{12}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{13}	1.00E-02	8.00E-02	5.00E-02	4.67E-02	2.60E-02	0.00E+00
f_{14}	1.00E-02	5.00E-02	3.00E-02	2.82E-02	2.13E-02	0.00E+00
f_{15}	0.00E+00	4.70E-01	2.20E-01	1.99E-01	1.55E-01	1.96E-02
f_{16}	4.00E-02	1.50E-01	8.00E-02	8.47E-02	4.96E-02	0.00E+00
f_{17}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{18}	0.00E+00	2.00E-02	1.00E-02	7.65E-03	6.32E-03	3.14E-01
f_{19}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{20}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{21}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{22}	0.00E+00	3.00E-02	0.00E+00	3.73E-03	2.76E-02	7.65E-01
f_{23}	0.00E+00	1.00E+02	0.00E+00	2.55E+01	5.10E+01	7.45E-01
f_{24}	0.00E+00	6.90E-01	0.00E+00	2.61E-02	1.33E-01	9.61E-01
f_{25}	1.00E+02	1.00E+02	1.00E+02	1.00E+02	0.00E+00	0.00E+00
f_{26}	8.00E-02	1.00E+02	5.29E+01	5.20E+01	3.19E+01	0.00E+00
f_{27}	2.50E-01	9.10E-01	5.40E-01	5.60E-01	2.92E-01	0.00E+00
f_{28}	0.00E+00	3.57E+02	3.43E+02	2.76E+02	1.60E+02	1.96E-01
f_{29}	1.00E+02	1.00E+02	1.00E+02	1.00E+02	0.00E+00	0.00E+00
f_{30}	1.84E+02	1.84E+02	1.84E+02	1.84E+02	3.25E-02	0.00E+00

consideraron MOEAs de distintos grupos. Particularmente, la validación experimental se desarrollo incluyendo a los algoritmos *Algoritmo Genético basado en Ordenación de los No-Dominados* (Non-Dominated Sorting Genetic Algorithm - NSGA-II) [15], *el MOEA basado en descomposición* (the MOEA based on Decomposition - MOEA/D) [54] y el *Algoritmo Evolutivo Multi-objetivo basado en la Métrica-S* (the *S*-Metric Selection Evolutionary Multi-objective Optimization Algorithm - SMS-EMOA) [4].) Estos algoritmos son representativos de los basados en dominancia, basados en descomposición y basados en indicadores respectivamente. Las siguientes sub-secciones describen cada uno de los paradigmas involucrados en los métodos seleccionados.

Cuadro 4: Resultados del DE-EDM con los problemas del CEC 2017

	Mejor	Peor	Mediana	Media	Sd	Razón de éxito
f_1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_5	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_6	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_9	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{10}	0.00E+00	1.20E-01	0.00E+00	1.65E-02	3.39E-02	7.45E-01
f_{11}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{12}	0.00E+00	2.20E-01	0.00E+00	6.37E-02	1.76E-01	6.67E-01
f_{13}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{14}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{15}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{16}	0.00E+00	2.10E-01	0.00E+00	2.47E-02	7.27E-02	8.82E-01
f_{17}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{18}	0.00E+00	1.00E-02	0.00E+00	1.96E-03	4.47E-03	8.04E-01
f_{19}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{20}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{21}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{22}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{23}	0.00E+00	3.00E+02	0.00E+00	3.49E+01	1.03E+02	8.82E-01
f_{24}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{25}	0.00E+00	1.00E+02	0.00E+00	3.92E+00	2.00E+01	9.61E-01
f_{26}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{27}	0.00E+00	3.87E+02	3.87E+02	2.05E+02	2.68E+02	1.96E-02
f_{28}	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E+00
f_{29}	1.45E+02	2.26E+02	2.18E+02	1.99E+02	4.21E+01	0.00E+00
f_{30}	3.95E+02	3.95E+02	3.95E+02	3.95E+02	2.10E-01	0.00E+00

4.1.1. Algoritmos Basados en el concepto de Dominancia - NSGA-II

Uno de los paradigmas mas reconocidos son los enfoques basados en dominancia. Los MOEAs que pertenecen a esta categoría se basan en la aplicación de la relación de dominancia para diseñar los distintos componentes de los mismos, especialmente en la fase de selección. Dado que la relación de dominancia no promueve la diversidad de forma implícita en el espacio de los objetivos se han desarrollado técnicas para obtener una diversidad en el espacio de los objetivos como es el niching, crowding y/o clustering que usualmente se integran. Una debilidad importante en los métodos basados en la relación de dominancia es la escalabilidad en términos de dimensionalidad en el espacio de los objetivos. De hecho la presión de selección es substancialmente reducida conforme el número de objetivos incrementan. A pesar de que se han desarrollado algunas estrategias para solventar este inconveniente [24], este parece ser la debilidad mas importante de este tipo de algoritmos.

El NSGA-II es uno de los algoritmos mas importantes de este grupo. Este algoritmo [15] considera un operador de selección especial el cual está basado en los procedimientos de la ordenación de las soluciones no dominadas y en el amontonamiento. El procedimiento que realiza la ordenación de las soluciones no dominadas es utiliza para proporcionas convergencia hacia el frente de Pareto mientras que el procedimiento de amontonamiento se utiliza para promover la diversidad en el espacio de los objetivos.

4.1.2. Algoritmos multi-objetivo basados en descomposición - MOEA/D

Los MOEAs basados en descomposición [54] transforman un MOP en un conjunto de problemas de optimización mono-objetivo de forma simultánea. Esta transformación puede ser lograda a través de distintos enfoques. Uno de los mas populares es por medio de la función pesada de Tchebycheff la cual requiere un conjunto de pesos bien distribuídos con el propósito de alcanzar soluciones bien distribuídas a lo largo del frente de Pareto. Sin embargo este tipo de algoritmos poseen una desventaja importante y está relacionada con la geometría que posee el frente de Pareto. El MOEA [54] es un MOEA popular de los basados en descomposición. Este incluye varias características, tales como la descomposición de los problemas, la agregación de los pesos con los objetivos y las restricciones de emparejamiento que están basadas en la definición de vecindarios. El algoritmo MOEA/D-DE es considerado como una variante popular del MOEA/D, el cual utiliza operadores de DE [41] y el operador de mutación polinomial [22] en la fase de reemplazo. Adicionalmente, este algoritmo tiene dos mecanismos especiales para mantener la diversidad de la población [55].

4.1.3. Algoritmo multi-objetivo basados en indcadores - SMS-EMOA

En optimización multi-objetivo se han desarrollado varios indicadores de calidad con el propósito de comparar el rendimiento de los MOEAs. Debido a que estos indicadores miden la calidad de las aproximaciones alcanzadas por los MOEAs se propuso un paradigma basado en en la aplicación de estos indicadores. Particularmente, los indicadores reemplazan a la relación de dominancia de Pareto con el propósito de guiar el proceso de optimización. Principalmente, el hipervolumén es uno de los indicador de calidad ampliamente aceptados denominado por su relación completa de Pareto (Pareto-compliance) [26]. Una de la principales ventajas de estos algoritmos es que los indicadores normalmente consideran tanto la calidad de las soluciones como su diversidad, por lo tanto no se requieren mecanismos adicionales para preservar la diversidad.

Un algoritmo popular que utiliza que es basado en indicadores es el SMS-EMOA [4]. Este MOEA es considerado como un algoritmo híbrido ya que utiliza tanto indicadores como el concepto de la dominancia de Pareto. Escencialmente, este algoritmo integra el procedimiento para ordenar a las soluciones no dominadas con la métrica del hipervolumén. Por lo tanto el SMS-EMOA aplica el hipervolumén como estimados de densidad siendo una tarea computacionalmente pesado. Particularmente, la fase de reemplazo elimina al individuo que pertenece al frente con peor rango y cuya contribución al hipervolumén sea mínima. Debido al comportamiento prometedor del SMS-EMOA se ha considerado como parte de nuestra validación

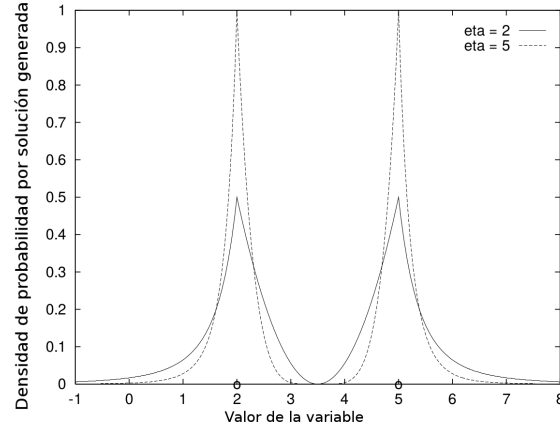


Figura 3: Probability density function of the SBX operator with indexes of distribution 2 and 5. The parents are located in 2 and 5 respectively.

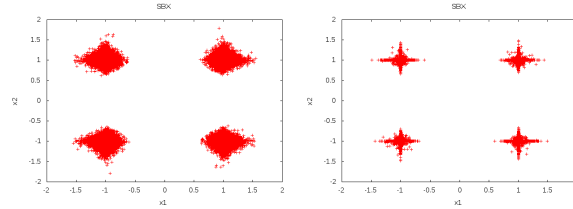


Figura 4: Simulations of the SBX operator with a distribution index equal to 20. Parents are located in $P_1 = (-1, 0, -1, 0)$ and $P_2 = (1, 0, 1, 0)$. The left simulation corresponds to a probability of altering a variable (δ_1 in Algorithm 3) equal to 1,0 and in the right it corresponds to 0,1.

experimental.

4.2. Operadores de cruce

Los operadores de cruce son diseñados para generar soluciones hijo utilizando la información de las soluciones padre. Estos combinan las características de dos o más soluciones padre con el propósito de generar nuevas soluciones candidatas. En base a que en la literatura existen varios operadores de cruce se han propuesto varias taxonomías para clasificarlos. Particularmente, las taxonomías se basan en varias características tales como la ubicación de las nuevas soluciones generadas o por el tipo de relaciones que existen en las variables.

Una taxonomía popular realiza la clasificación de los operadores de cruce en basados en variables o basados en vectores. En la categoría donde son basados en las variables, cada variable de las soluciones padre son combinadas para crear nuevos valores, esto se realiza de forma independiente y en base a una probabilidad especificada con anterioridad Este tipo de operadores son ideales para lidiar con problemas separables. Algunos operadores que pertenecen a esta categoría son el *Operador de Cruce Ciego* (the Blend

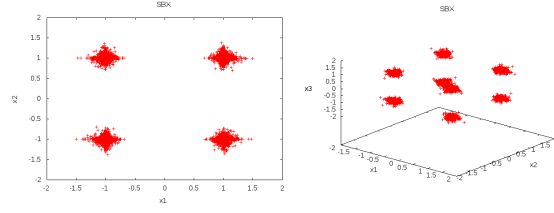


Figura 5: Simulations of the SBX operator with a distribution index equal to 20. Parents are located in $P_1 = (-1, 0, -1, 0)$ and $P_2 = (1, 0, 1, 0)$ and $P_1 = (-1, 0, -1, 0, -1, 0)$ and $P_2 = (1, 0, 1, 0, 1, 0)$ for two and three variables respectively.

Crossover - BLX) [20] y el SBX [1]. De forma alternativa, los operadores de recombinación basados en vectores son diseñados para considerar la dependencia que existe entre las variables. Este tipo de operadores regularmente realizan una combinación lineal de las soluciones involucradas. Algunos operadores que pertenecen a esta categoría son *El Operador de Cruce Unimodal Normalmente Distribuido* (The Unimodal Normally Distributed Crossover - UNDX) [38], y *El Operador de Cruce basado en el Simplex* (The simplex crossover - SPX) [47]. Adicionalmente, los operadores de cruce pueden ser clasificados como basados en los padres y basados en la media [27]. En los operadores basados en los padres, las soluciones hijo son creadas alrededor de cada solución padre, mientras que en los operadores basados en la media existen una tendencia de crear a las soluciones hijo alrededor de la media generada por las soluciones padres. Entre los operadores de cruce el SBX es probablemente uno de los más utilizados, por lo tanto esta sección se centra en este operador de cruce.

4.2.1. El Operador de Cruce Basado en Simulación Binaria - SBX

Los operadores de reproducción son uno de los componentes mas relevantes para influencias el proceso de búsqueda en los EAs. Específicamente, los operadores de cruce y mutación están altamente relacionados con la diversidad de las soluciones. Por lo tanto, los operadores considerados afectan la calidad de las soluciones de forma significativa.

Probablemente *El Operador de Cruce Basado en Simulación Binaria* (Simulated Binary Crossover - SBX) [13] es uno de los operadores mas populares en dominios continuos y por lo tanto ha sido utilizado de forma extensiva en muchos MOEAs [15, 4]. El operador SBX es clasificado con un operador basado en los padres, lo cual significa las soluciones que corresponden a los hijos (c_1 and c_2) serán creadas alrededor de los valores de los padres (p_1 and p_2). Específicamente, el proceso para generar los valores de las soluciones hijo se basa en una distribución de probabilidad. Esta distribución controla el factor de dispersión $\beta = |c_1 - c_2|/|p_1 - p_2|$ el cual es definido como la razón entre la dispersión de los valores de las soluciones hijo y los valores de las soluciones padre. En orden, esta función de densidad se define en base a un índice de distribución η_c (es un parámetro de control especificado por el usuario) el cual altera la capacidad de exploración. Específicamente, un índice pequeño induce una probabilidad elevada de crear valores de las

soluciones hijo alejadas de los valores de las soluciones padre. mientras que índices elevados tienden a crear soluciones muy similares a las soluciones padre, esto se demuestra en la Figura 3.

Para crear una solución hijo se utiliza una distribución de probabilidad que es definida en función de $\beta \in [0, \infty]$ de la siguiente forma:

$$P(\beta) = \begin{cases} 0,5(\eta_c + 1)\beta^{\eta_c}, & \text{si } \beta \leq 1 \\ 0,5(\eta_c + 1)\frac{1}{\beta^{\eta_c+2}}, & \text{de otra forma} \end{cases} \quad (8)$$

Basado en la propiedad para preservar la media de los valores que corresponden a las soluciones hijo y padre, el SBX tiene las siguientes propiedades:

- Los dos valores de las soluciones hijo son equidistantes de los valores de las soluciones padre.
- Existe una probabilidad no nula de crear soluciones hijo en el espacio factible entero por cualquier par de soluciones padre.
- La probabilidad general de crear un par de soluciones hijo dentro del rango de las soluciones padre es idéntico a la probabilidad general de crear un par de soluciones hijo afuera del rango de las soluciones padre.

Por lo tanto, al considerar dos valores padre (p_1 and p_2), pueden ser creados dos valores hijo (c_1 and c_2) como una combinación lineal de los valores padre con un número aleatorio uniforme $u \in [0, 1]$ de la siguiente forma:

$$\begin{aligned} c_1 &= 0,5(1 + \beta(u))p_1 + 0,5(1 - \beta(u))p_2 \\ c_2 &= 0,5(1 - \beta(u))p_1 + 0,5(1 + \beta(u))p_2 \end{aligned} \quad (9)$$

El parámetro $\beta(u)$ depende en el número aleatorio u de la siguiente forma:

$$\beta(u) = \begin{cases} (2u)^{\frac{1}{\eta_c+1}}, & \text{si } u \leq 0,5, \\ (\frac{1}{2(1-u)})^{\frac{1}{\eta_c+1}}, & \text{de otra forma} \end{cases} \quad (10)$$

La ecuación anterior es formulada en base a un problema de optimización sin límites en las variables. Sin embargo en problemas prácticos cada variable es limitada dentro de un límite inferior y superior. Por lo tanto, para considerar los límites del espacio de decisión [14] propusieron una modificación de la distribución de probabilidad en la ecuación (11). Es importante resaltar que esta última variante es popularmente utilizada.

$$\beta(u) = \begin{cases} (2u(1-\gamma))^{\frac{1}{\eta_c+1}}, & \text{si } u \leq 0,5/(1-\gamma), \\ (\frac{1}{2(1-u(1-\gamma))})^{\frac{1}{\eta_c+1}}, & \text{de otra forma} \end{cases} \quad (11)$$

Algorithm 3 Operador de Cruce basado en Simulación Binaria (SBX)

```
1: Entrada: Soluciones padre ( $P_1, P_2$ ), Índica de distribución ( $\eta_c$ ), Probabilidad de cruce ( $P_c$ ).
2: Salida: Soluciones hijo ( $C_1, C_2$ ).
3: if  $U[0, 1] \leq P_c$  then
4:   for para cada variable  $d$  do
5:     if  $U[0, 1] \leq \delta_1$  then
6:       Generar  $C_{1,d}$  utilizando las ecuaciones (11) y (12).
7:       Generar  $C_{2,d}$  utilizando las ecuaciones (11) y (13).
8:     if  $U[0, 1] \leq (1 - \delta_2)$  then
9:       Intercambiar  $C_{1,d}$  con  $C_{2,d}$ .
10:    else
11:       $C_{1,d} = P_{1,d}$ .
12:       $C_{2,d} = P_{2,d}$ .
13: else
14:    $C_1 = P_1$ .
15:    $C_2 = P_2$ .
```

$$c_1 = 0,5(1 + \beta(u))p_1 + 0,5(1 - \beta(u))p_2 \quad (12)$$

$$c_2 = 0,5(1 + \beta(u))p_1 + 0,5(1 - \beta(u))p_2 \quad (13)$$

De esta forma en base a la ecuación (12) En este caso se calcula al valor del padre p_1 con el valor hijo c_1 mas cercano. Considerando que $p_1 < p_2$ y con un límite inferior igual a a , se tiene que $\gamma = 1/(\alpha^{\eta_c+1})$, donde $\alpha = 1 + (p_1 - a)/(p_2 - p_1)$. Similarmente, el segundo valor hijo c_2 es calculado con $\alpha = 1 + (b - p_2)/(p_2 - p_1)$, donde b corresponde al límite superior. Entonces, el segundo valor hijo es calculado como se indica en la ecuación (13).

Es importante hacer mención de en base a lo reportado en [1] se han proporcionaron varias extensiones del SBX. Los autores consideraron una estrategia simple para escoger las variables que se van a cruzar [38]. Específicamente, en base a los principios del operador de cruce uniforme cada variable es cruzada con una probabilidad del 0,5. Sin embargo, los autores reconocieron las implicaciones que existen con los problemas donde un grado elevado de dependencia puede existir entre las variables. En cualquier caso, actualmente esta es la forma mas común de aplicar el SBX en problemas con múltiple variables.

4.2.2. Implementación y análisis del operador SBX

En este apartado se discuten algunas de las principales características de la implementación más utilizada del operador SBX, los cuales son utilizados con problemas de múltiples variables. Escencialmente, se consideran tres componentes clave los cuales afectar el rendimiento de los MOEAs. Inicialmente, como ya se mencionó con anterioridad cada variable es alterada dada una probabilidad de 0,5. Si este valor de probabilidad es incrementado, entonces existe una tendencia de generar valores hijo mas distante de los padres debido a que mas variables son modificadas por cada operación. De acuerdo a esto, sería adecuado modificar una variable en los problemas separables. Sin embargo, parece ser mas conveniente modificar un conjunto

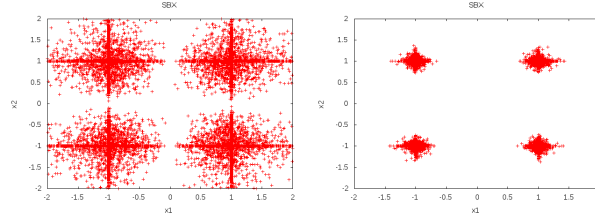


Figura 6: Simulation of the SBX operator sampling 10,000 children values, the parents are located in $P_1 = (-1,0,-1,0)$ and $P_2 = (1,0,1,0)$. The left and right are with a distribution index of 2 and 20 respectively.

de variables de forma simultánea en problemas no-separables. En la figura 4 se pueden observar las implicaciones de variar esta probabilidad, donde se considera un problema con dos variables. Particularmente, en la parte derecha se muestra que una probabilidad pequeña provoca una tendencia de exploración donde algunas variables no son modificadas, es decir hay una tendencia de generar desplazamientos paralelos a los ejes. Esta característica es ideal para problemas separables. De forma alternativa, en la parte izquierda se muestra que utilizando una probabilidad elevada existe un comportamiento de búsqueda distinto donde la tendencia anterior desaparece, lo cual podría ser ideal para problemas no separables. Es importante destacar que existe una relación entre esta probabilidad y con el índice de distribución, de hecho estos dos factores tienen un efecto directo en la similitud que exist entre las soluciones padre e hijo.

El segundo aspecto clave es después de generar dos valores de las soluciones hijo con la distribución SBX, ya que estos valores son intercambiados con una probabilidad fija que usualmente es 0,5, es decir el valor de la solución hijo c_1 no siempre es heredado por la solución padre mas cercana p_1 . Esta es una característica no muy discutida, sin embargo es un aspecto muy relevante y afecta al rendimiento del algoritmo. En algunos contextos esta probabilidad se identifica como “Probabilidad de cruce uniforme por variable” (Variable uniform crossover probability) [48] o “Recombinación Discreta” (Discrete Recombination) [36].

Desde que en el ámbito multi-objetivo se promueve más diversidad en las variables de decision de forma implícita, estos intercambios podrían ser altamente disruptivos. De hecho, en algún sentido y dado a esto, no es totalmente claro que el SBX pueda ser categorizado como un operador totalmente basado en los padres. Estos intercambios que existen entre los valores de las soluciones hijo tienen un efecto de realizar múltiples “reflexiones” en el espacio de búsqueda. De hecho, conforme incrementa la dimensionalidad en el espacio de las variables, el número de regiones cubiertas incrementa de forma exponencial como se puede observar en los casos de dos y tres dimensiones en la figura 5. Es importante notar que esta característica tiene un efecto relevante en la distancia entre las soluciones padre y las soluciones hijo.

Finalmente, siendo quizás la característica mas conocida del operador SBX es el índice de distribución. Un índice de distribución pequeño provoca un grado de exploración elevado. De hecho, un índice de distribución de la unidad tiene un efecto similar al *Operador de Recombinación Difusa* (Fuzzy Recombination Operator) [50]. En la figura 6 se puede observar el efecto de aplicar distintos índices de distribución. Particularmente, en

la parte izquierda se considera un índice de distribución pequeño, mientras en la parte derecha se considera un índice de distribución grande, se observa que este último genera soluciones candidatas similares a las soluciones padre.

En el algoritmo 3 se muestra la implementación del operador SBX. Este pseudocódigo está basado en la implementación que está integrado en el código NSGA-IIpropuesto por Deb y otros [15], el cual es considerada como la variante mas popular. Como parámetros de entrada se requieren dos soluciones padre (P_1 and P_2), y como salida se obtienen dos soluciones hijo (C_1 and C_2). El primero y el segundo componente que mencionados previamente corresponden a las líneas 5 y 9 respectivamente. Como es usual, el caso clásico del operador SBX es configurado asignando los parámetros $\delta_1 = \delta_2 = 0,5$ y $\eta_c = 20$. Es important notar que la implementación clásica no considera la dimensión de la variable o el criterio de paro en sus parámetros internos.

4.3. Propuesta

Basado en el análisis anterior y con el propósito de inducir un balance entre exploración e intensificación, se proponen las siguiente modificaciones. Primeramente, se modifica la probabilidad de alterar una variable (δ_1) durante la ejecución de forma dinámica. La intención de esta modificación es incrementar la capacidad de exploración en las primeras etapas, esto alterando de forma simultánea un conjunto de variables y conforme la ejecución procede se reduce el número de variables que son modificadas. El valor de δ_1 se cambia en base a un modelo lineal decreciente, donde inicialmente es fijado a 1,0 y entonces es decrementado hasta la mitad del total de generaciones con un valor de 0,5. Esta último valor es mantenido hasta el final de la ejecución, es decir desde la mitad de la ejecución este parámetro se comporta similar a la implementación tradicional del SBX.

Para asignar el valor δ_1 se utiliza la ecuación (14), donde $G_{Transcurridas}$ corresponde a la generación actual y G_{Total} corresponde al número total de generaciones.

Similarmente, el segundo cambio está relacionado con la probabilidad de aplicar reflexiones ($1 - \delta_2$). En este caso δ_2 es actualizado de acuerdo a la ecuación (14), por lo tanto la probabilidad de aplicar una reflexión incrementa de 0,0 a 0,5 durante la ejecución. Esta modificación se realiza con el propósito de evitar el comportamiento disruptivo de intercambiar variables en las primeras generaciones ya que esto provocaría modificaciones muy drásticas. De esta forma, sería mas sensato aplicar estas reflexiones una vez que los individuos convergen a cierto grado- Por lo tanto, esta probabilidad es incrementado a 0,5, siendo el valor utilizando en la implementación del SBX estándar.

$$\delta_1 = \delta_2 = \max \left(0,5, 1,0 - \frac{G_{Transcurridas}}{G_{Total}} \right) \quad (14)$$

Finalmente, el índice de distribución también es modificado durante la ejecución. De esta forma, en la primeras etapas se promueve un índice de distribución pequeño con el propósito de incrementar la capacidad

Cuadro 5: Puntos de referencias para el indicador HV

Instancias	Punto de referencia
WFG1-WFG9	$[2, 1, \dots, 2m + 0, 1]$
DTLZ 1, 2, 4	$[1, 1, \dots, 1, 1]$
DTLZ 3, 5, 6	$[3, \dots, 3]$
DTLZ7	$[1, 1, \dots, 1, 1, 2m]$
UF 1-10	$[2, \dots, 2]$

Cuadro 6: Información estadística de las métricas considerando dos objetivos

	NSGA-II						MOEA/D						SMS-EMOA					
	1	2	3	4	5	DE	1	2	3	4	5	DE	1	2	3	4	5	DE
Average HV	0.88	0.90	0.90	0.91	0.93	0.94	0.87	0.87	0.87	0.90	0.91	0.91	0.88	0.89	0.87	0.91	0.92	0.93
Average IGD+	0.12	0.09	0.11	0.07	0.06	0.05	0.14	0.12	0.14	0.09	0.08	0.07	0.13	0.11	0.14	0.08	0.07	0.05

de exploración del SBX. Posteriormente es decrementado de forma lineal, lo cual tiene implica que la curva de distribución se cierre, por lo tanto se promueve un mayor grado de intensificación en las etapas finales. El incremento lineal es indicado en la ecuación (15), por lo tanto el índice de distribución es alterado de 2 a 22. Es importante hacer mención que ya se han considerado modificaciones similares al índice de distribución [57], [22].

$$\eta_c = 2 + 20 \times \left(\frac{G_{Elapsed}}{G_{End}} \right) \quad (15)$$

4.4. Resultados

En este apartado se analizan los resultados obtenidos con las variantes dinámicas del SBX (DSBX). Cada uno de los casos estudiados y una propuesta final se integraron en los algoritmos NSGA-II, MOEA/D and SMS-EMOA. Primeramente, es analizado cada uno de los tres componentes previamente analizados. Posteriormente se construye un caso donde se consideran dos componentes de forma simultánea. Se consideran los problemas de prueba WFG [25], DTLZ [16] and UF [55]. Con el propósito de comparar nuestra extensión del SBX con otros operadores populares en la validación experimental incluimos una variante de evolución diferencial mejor conocida como DEMO [48]. Debido a que todos los algoritmos con estocástico cada ejecución se repitió 35 veces con distintas semilla La configuración global que es aplicada a todos los algoritmos es como es indica a continuación. Se asigna el criterio de paro a 25,000 generaciones, el tamaño de la población a 100, se configuraron los problemas de prueba WFG con dos y tres objetivos, además se consideraron 24 variables, donde 20 variables eran considerados como parámetros de distancia y 4 se consideraron como parámetros de posición. Como es sugerido en [16] se consideran $n = M + r - 1$ variables de decisión en los problemas de prueba DTLZ, donde para los problemas DTLZ1, DTLZ2 - DTLZ6 y DTLZ7 se consideraron $r = \{5, 10, 20\}$ respectivamente. Para los problemas de prueba UF se asignaron 10 variable de decisión. Finalmente, se asignó al operador de mutación polinomial con una probabilidad de cruce de $1/n$ y

Cuadro 7: Información estadística de las métricas considerando tres objetivos

	NSGA-II						MOEA/D						SMS-EMOA					
	1	2	3	4	5	DE	1	2	3	4	5	DE	1	2	3	4	5	DE
Average HV	0.87	0.84	0.87	0.87	0.87	0.85	0.84	0.84	0.84	0.86	0.86	0.85	0.90	0.89	0.88	0.91	0.91	0.91
Average IGD+	0.13	0.16	0.13	0.12	0.12	0.13	0.15	0.14	0.15	0.11	0.11	0.13	0.11	0.11	0.13	0.09	0.09	0.13

con un índice de distribución de 50, mientras que el operador de cruce SBX se asignó con una probabilidad de cruce de 0,9 y un índice de distribución de 20. A continuación se especifica la parametrización adicional de cada algoritmo:

- **DEMO**: $CR = 0.3$ and $F = 0.5$.
- **SMS-EMOA**: desplazamiento para calcular el $HV = 100$.
- **MOEA/D**: tamaño de la vecindad = 10, el número de actualizaciones por subproblema (nr) = 2 y $\delta = 0,9$. i

Para comparar los frentes obtenidos de cada método se utiliza el hipervolumen normalizado (HV) y la *Distancia Generacional Invertida Modificada* (Inverted Generational Distance Plus - IGD+). En la tabla 5 se presentan los puntos de referencia utilizados para el indicador del hipervolumen los cuales son similares a los utilizados en [3, 56]. Por otra parte para comparar los resultados estadísticamente (valores del IGD+ y HV), se siguió un procedimiento similar al que se propuso en [17]. En primer lugar se utilizó el test Shapiro-Wilk para comprobar si los resultados se ajustaban a una distribución Gaussiana. Por otro lado, para los casos que no se ajustaban a distribuciones Gaussianas, se utilizó el test de Kruskal-Wallis. En todos los casos se fijó el nivel de confianza al 95 %. Se considera que un algoritmo X es superior a un algoritmo Y , si el procedimiento anterior reporta diferencias significativas y si la media y mediana del hipervolumen obtenido por el método X son superiores a las obtenidas por el método Y .

4.5. Análisis de cada componente en el operador SBX

En este apartado se analiza el efecto que cada componente tiene de forma independiente al ser dinámicamente modificado. Basados en el algoritmo 3 el efecto de cada componente es analizado a través de cuatro casos. Cada caso es descrito a continuación.

- **Caso 1**: Se aplica la versión estándar del operador SBX donde $\delta_1 = \delta_2 = 0,5$ y $\eta_c = 20$.
- **Caso 2**: Se actualiza el valor δ_1 como se indica en la ecuación (14), $\delta_2 = 0,5$ y $\eta_c = 20$.
- **Caso 3**: Se actualiza el valor δ_2 como se indica en la ecuación (14), $\delta_1 = 0,5$ y $\eta_c = 20$.
- **Caso 4**: Se actualiza el índice de distribución de acuerdo a (15), $\delta_1 = \delta_2 = 0,5$.

Cuadro 8: Resumen de las pruebas estadísticas

NSGA-II															
	1			2			3			4			5		
	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔
HV-2obj	16	29	47	6	61	25	28	19	45	31	23	38	54	3	35
HV-3obj	15	19	42	12	50	14	17	15	44	33	10	33	26	9	41
IGD-2obj	14	30	48	4	60	28	25	17	50	33	19	40	52	2	38
IGD-3obj	14	18	44	13	44	19	18	15	43	33	15	28	23	9	44

MOEA/D															
	1			2			3			4			5		
	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔
HV-2obj	15	33	44	10	60	22	25	26	41	39	18	35	57	9	26
HV-3obj	10	22	44	12	39	25	11	19	46	24	10	42	38	5	33
IGD-2obj	16	31	45	9	60	23	23	27	42	37	17	38	57	7	28
IGD-3obj	12	22	42	13	43	20	13	24	39	30	9	37	40	10	26

SMS-EMOA															
	1			2			3			4			5		
	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔	↑	↓	↔
HV-2obj	9	35	48	7	43	42	16	31	45	41	9	42	53	8	31
HV-3obj	7	21	48	9	35	32	13	21	42	27	6	43	31	4	41
IGD-2obj	10	34	48	15	48	29	12	33	47	41	12	39	55	6	31
IGD-3obj	8	20	48	13	30	33	9	19	48	22	5	49	27	5	44

En las tablas 6 y 7 se muestra información del HV normalizado [57] y del IGD+ [26] para cada caso (Posteriormente se analiza el Caso 5). Específicamente, se muestra la media del HV e IGD+ para todos los problemas en dos y tres objetivos. Se observa que considerando dos y tres objetivos el Caso 4 mejora al Caso 1, al Caso 2 y al Caso 3 en todos los algoritmos. Por lo tanto se aprecian los beneficios de incrementar el índice de distribución durante la ejecución. Esto sucede porque inicialmente la curva que corresponde a la distribución de probabilidad conduce a un grado elevado de exploración, mientras que al transcurrir las generaciones se cambia gradualmente a un procedo de intensificación. Por otra parte, al considerar tres objetivos el Caso 2 presentó un rendimiento menor que al Caso 1. Por lo tanto, se observa que alterar todas las variables provoca un comportamiento muy disruptivo. Quizás los resultados podrían mejorar si el parámetro δ_1 es alterado de una forma distinta, sin embargo esto se deja como trabajo futuro. Los análisis anteriores únicamente consideran la media obtenida en todos los problemas de prueba. Sin embargo, dependiendo en

el tipo de problema el rendimiento del algoritmo podría cambiar. Esto se analiza en la siguiente sección. Adicionalmente, se anexan resultados detallados².

4.6. Modificación simultánea de varios componentes

En base a los resultados obtenidos con anterioridad se propone una variante del operador SBX. En esta variante se considera el Caso 3 y el Caso 4 de forma simultánea, es decir existen un cambio dinámico en el parámetro δ_2 y en el índice de distribución. El Caso 2 donde se aplica un mecanismo para actualiza δ_1 no se considera debido que los beneficios con esta variante no fueron significativos. Particularmente, el Caso 5 es construido en base al Algoritmo 3 y es configurado como se indica a continuación. El parámetro δ_1 es asignado con 0,5 ya que es la forma de la versión estándar del SBX. En base al Caso 3, el parámetro δ_2 es actualizado de acuerdo a la ecuación (14). Finalmente, en base al Caso 4, el parámetro δ_2 es actualizado como se indica en la ecuación (15).

De acuerdo a la media del HV y del IGD+ que se obtuvieron en el Caso 5 (revisar tablas 6 and 7), se pueden observar los beneficios de integrando los Casos 3 y 4. En el caso de dos objetivos se observa una ventaja significativa, sin embargo en el caso de tres objetivo los Casos 4 y 5 son muy similares en base a la media. Además, al considerar tres objetivos los resultados que se obtuvieron en el Caso 5 son superior a los obtenidos con DE, mientras que al considerar la versión estándar de SBX se observa un deterioro. Por lo tanto, el operador SBX puede generar resultados similares o superiores que los algoritmos DEMO si éste primero es configurado de una forma adecuada.

Finalmente, debido a que los análisis previos únicamente consideran la media entre la problemas de prueba, a continuación se presenta un análisis adicional para comprender mejor las contribuciones de los distintos casos. Particularmente, se realizan comparaciones en pares en base a pruebas estadísticas entre los cinco casos donde es considerado el SBX y DSBX. Este procedimiento se realizó de forma independiente para cada el NSGA-II, MOEA/D y el SMS-EMOA. En la tabla 8, se muestra los resultados de las pruebas estadísticas. Para cada algoritmo y para cada caso, la columna “ \uparrow ” reporta el número de comparaciones donde las pruebas estadísticas confirmaron la superioridad del caso correspondiente, mientras que en la columna “ \downarrow ” se reporta el número de veces donde este caso fue inferior y la columna “ \longleftrightarrow ” indica el número de comparaciones donde la diferencia estadística no fue significativamente distinto. Los beneficios obtenidos por el caso 5 son suficientemente claros. Únicamente el Caso 4 obtuvo mejores resultados que el Caso 5 al considerar tres objetivos y al algoritmo NSGA-II. Entonces, los resultados pueden mejorar aún mas al integrar varias modificaciones dinámicas. Mas aún, los resultados que confirman varias ventajas al comparar nuestras variantes con la versión estándar del operador SBX (Caso 1). El único caso que no es superiormente claro a la versión estándar del SBX es el Caso 2, el cual fue discutido con anterioridad.

²<https://github.com/joelchaconcastillo/SBX-CEC2018>.

Referencias

- [1] R. B. Agrawal, K. Deb, K. Deb, R. B. Agrawal, Simulated binary crossover for continuous search space, Tech. rep. (1994).
- [2] N. H. Awad, M. Z. Ali, P. N. Suganthan, R. G. Reynolds, An ensemble sinusoidal parameter adaptation incorporated with l-shade for solving cec2014 benchmark problems, in: *Evolutionary Computation (CEC)*, 2016 IEEE Congress on, IEEE, 2016, pp. 2958–2965.
- [3] J. A. M. Berenguer, C. A. C. Coello, Evolutionary many-objective optimization based on kuhn-munkres' algorithm, in: *International Conference on Evolutionary Multi-Criterion Optimization*, Springer, 2015, pp. 3–17.
- [4] N. Beume, B. Naujoks, M. Emmerich, SMS-EMOA: Multiobjective selection based on dominated hypervolume, *European Journal of Operational Research* 181 (3) (2007) 1653 – 1669.
- [5] T. Blickle, L. Thiele, A comparison of selection schemes used in evolutionary algorithms, *Evolutionary Computation* 4 (4) (1996) 361–394.
- [6] A. Bolufé-Röhler, S. Estévez-Velarde, A. Piad-Morffis, S. Chen, J. Montgomery, Differential evolution with threshold convergence, in: *Evolutionary Computation (CEC)*, 2013 IEEE Congress on, IEEE, 2013, pp. 40–47.
- [7] J. Brest, M. S. Maučec, B. Bošković, il-shade: Improved l-shade algorithm for single objective real-parameter optimization, in: *Evolutionary Computation (CEC)*, 2016 IEEE Congress on, IEEE, 2016, pp. 1188–1195.
- [8] J. Brest, M. S. Maučec, B. Bošković, Single objective real-parameter optimization: Algorithm jso, in: *Evolutionary Computation (CEC)*, 2017 IEEE Congress on, IEEE, 2017, pp. 1311–1318.
- [9] U. K. Chakraborty, *Advances in differential evolution*, vol. 143, Springer, 2008.
- [10] S. Das, P. N. Suganthan, Differential evolution: A survey of the state-of-the-art, *IEEE transactions on evolutionary computation* 15 (1) (2011) 4–31.
- [11] D. Dasgupta, Z. Michalewicz, *Evolutionary algorithms in engineering applications*, Springer Science & Business Media, 2013.
- [12] K. A. De Jong, *Evolutionary computation: a unified approach*, MIT press, 2006.
- [13] K. Deb, R. B. Agrawal, Simulated binary crossover for continuous search space, *Complex Systems* 9 (3) (1994) 1–15.
- [14] K. Deb, H.-G. Beyer, Self-adaptive genetic algorithms with simulated binary crossover, *Secretary of the SFB 531*, 1999.
- [15] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation* 6 (2) (2002) 182–197.
- [16] K. Deb, L. Thiele, M. Laumanns, E. Zitzler, *Scalable test problems for evolutionary multiobjective optimization*, Springer, 2005.
- [17] J. J. Durillo, A. J. Nebro, C. A. C. Coello, J. Garcia-Nieto, F. Luna, E. Alba, A Study of Multiobjective Metaheuristics When Solving Parameter Scalable Problems, *IEEE Transactions on Evolutionary Computation* 14 (4) (2010) 618–635.
- [18] A. E. Eiben, C. A. Schippers, On evolutionary exploration and exploitation, *Fundamenta Informaticae* 35 (1-4) (1998) 35–50.
- [19] S. Elsayed, N. Hamza, R. Sarker, Testing united multi-operator evolutionary algorithms-ii on single objective optimization problems, in: *Evolutionary Computation (CEC)*, 2016 IEEE Congress on, IEEE, 2016, pp. 2966–2973.
- [20] L. J. Eshelman, Real coded genetic algorithms and interval-schemata, *Foundations of genetic algorithms* 2 (1993) 187–202.
- [21] F. Glover, G. Kochenberger, *Handbook of metaheuristics (international series in operations research and management science)*, JOURNAL-OPERATIONAL RESEARCH SOCIETY 56 (5) (2005) 614–614.
- [22] M. Hamdan, The distribution index in polynomial mutation for evolutionary multiobjective optimisation algorithms: An experimental study, in: *International Conference on Electronics Computer Technology*, Kanyakumari, India, 2012.
- [23] F. Herrera, M. Lozano, Adaptation of genetic algorithm parameters based on fuzzy logic controllers, *Genetic Algorithms and Soft Computing* 8 (1996) 95–125.
- [24] C. Horoba, F. Neumann, Benefits and drawbacks for the use of epsilon-dominance in evolutionary multi-objective op-

- timization, in: Proceedings of the 10th annual conference on Genetic and evolutionary computation, ACM, 2008, pp. 641–648.
- [25] S. Huband, L. Barone, L. While, P. Hingston, A Scalable Multi-objective Test Problem Toolkit, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 280–295.
- [26] H. Ishibuchi, H. Masuda, Y. Tanigaki, Y. Nojima, Modified distance calculation in generational distance and inverted generational distance, in: International Conference on Evolutionary Multi-Criterion Optimization, Springer, 2015, pp. 110–125.
- [27] H. Jain, K. Deb, Parent to mean-centric self-adaptation in sbx operator for real-parameter optimization, Swarm, Evolutionary, and Memetic Computing (2011) 299–306.
- [28] A. Kumar, R. K. Misra, D. Singh, Improving the local search capability of effective butterfly optimizer using covariance matrix adapted retreat phase, in: Evolutionary Computation (CEC), 2017 IEEE Congress on, IEEE, 2017, pp. 1835–1842.
- [29] J. Lampinen, I. Zelinka, et al., On stagnation of the differential evolution algorithm, in: Proceedings of MENDEL, 2000, pp. 76–83.
- [30] M. Lozano, F. Herrera, J. R. Cano, Replacement strategies to preserve useful diversity in steady-state genetic algorithms, Information Sciences 178 (23) (2008) 4421–4433.
- [31] S. W. Mahfoud, Crowding and preselection revisited, Urbana 51 (1992) 61801.
- [32] D. Molina, F. Moreno-García, F. Herrera, Analysis among winners of different ieeec cec competitions on real-parameters optimization: Is there always improvement?, in: Evolutionary Computation (CEC), 2017 IEEE Congress on, IEEE, 2017, pp. 805–812.
- [33] J. Montgomery, Differential evolution: Difference vectors and movement in solution space, in: Evolutionary Computation, 2009. CEC’09. IEEE Congress on, IEEE, 2009, pp. 2833–2840.
- [34] J. Montgomery, S. Chen, An analysis of the operation of differential evolution at high and low crossover rates, in: Evolutionary Computation (CEC), 2010 IEEE Congress on, IEEE, 2010, pp. 1–8.
- [35] J. Montgomery, S. Chen, A simple strategy for maintaining diversity and reducing crowding in differential evolution, in: Evolutionary Computation (CEC), 2012 IEEE Congress on, IEEE, 2012, pp. 1–8.
- [36] H. Mühlenbein, D. Schlierkamp-Voosen, Predictive models for the breeder genetic algorithm i. continuous parameter optimization, Evolutionary computation 1 (1) (1993) 25–49.
- [37] J. A. Nelder, R. Mead, A simplex method for function minimization, The computer journal 7 (4) (1965) 308–313.
- [38] I. Ono, H. Kita, S. Kobayashi, Advances in evolutionary computing, chap. A Real-coded Genetic Algorithm Using the Unimodal Normal Distribution Crossover, Springer-Verlag New York, Inc., New York, NY, USA, 2003, pp. 213–237.
URL <http://dl.acm.org/citation.cfm?id=903758.903767>
- [39] H. M. Pandey, A. Chaudhary, D. Mehrotra, A comparative review of approaches to prevent premature convergence in ga, Applied Soft Computing 24 (2014) 1047–1077.
- [40] M. Pilát, Evolutionary multiobjective optimization: A short survey of the state-of-the-art, Proceedings of the Contributed Papers Part I-Mathematics and Computer Sciences, WDS, Prague, Czech (2010) 1–4.
- [41] K. Price, R. M. Storn, J. A. Lampinen, Differential evolution: a practical approach to global optimization, Springer Science & Business Media, 2006.
- [42] W. Price, Global optimization by controlled random search, Journal of Optimization Theory and Applications 40 (3) (1983) 333–348.
- [43] Â. A. Sá, A. O. Andrade, A. B. Soares, S. J. Nasuto, Exploration vs. exploitation in differential evolution, in: AISB 2008 Convention Communication, Interaction and Social Intelligence, vol. 1, 2008, p. 57.
- [44] C. Segura, C. A. C. Coello, E. Segredo, A. H. Aguirre, A novel diversity-based replacement strategy for evolutionary algorithms, IEEE transactions on cybernetics 46 (12) (2016) 3233–3246.

- [45] A. G. Slim Bechikh, Rituparna Datta, Recent Advances in Evolutionary Multi-objective Optimization, springer.
- [46] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization* 11 (4) (1997) 341–359.
- [47] R. Storn, K. Price, Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces, *J. of Global Optimization* 11 (4) (1997) 341–359.
URL <http://dx.doi.org/10.1023/A:1008202821328>
- [48] T. Tušar, B. Filipič, Differential evolution versus genetic algorithms in multiobjective optimization, in: *International Conference on Evolutionary Multi-Criterion Optimization*, Springer, 2007, pp. 257–271.
- [49] M. Črepinšek, S.-H. Liu, M. Mernik, Exploration and exploitation in evolutionary algorithms: A survey, *ACM Computing Surveys* 45 (3) (2013) 35:1–35:33.
- [50] H.-M. Voigt, H. Mühlenbein, D. Cvetkovic, Fuzzy recombination for the breeder genetic algorithm, in: *Proc. Sixth Int. Conf. on Genetic Algorithms*, Citeseer, 1995.
- [51] M. Yang, C. Li, Z. Cai, J. Guan, Differential evolution with auto-enhanced population diversity, *IEEE transactions on cybernetics* 45 (2) (2015) 302–315.
- [52] D. Zaharie, Control of population diversity and adaptation in differential evolution algorithms, in: *Proc. of MENDEL*, vol. 9, 2003, pp. 41–46.
- [53] J. Zhang, A. C. Sanderson, Jade: adaptive differential evolution with optional external archive, *IEEE Transactions on evolutionary computation* 13 (5) (2009) 945–958.
- [54] Q. Zhang, H. Li, MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition, *IEEE Transactions on Evolutionary Computation* 11 (6) (2007) 712–731.
- [55] Q. Zhang, W. Liu, H. Li, The performance of a new version of moea/d on cec09 unconstrained mop test instances, in: *Evolutionary Computation, 2009. CEC’09. IEEE Congress on*, IEEE, 2009, pp. 203–208.
- [56] Q. Zhu, Q. Lin, Z. Du, Z. Liang, W. Wang, Z. Zhu, J. Chen, P. Huang, Z. Ming, A novel adaptive hybrid crossover operator for multiobjective evolutionary algorithm , *Information Sciences* 345 (2016) 177 – 198.
- [57] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach, *IEEE transactions on Evolutionary Computation* 3 (4) (1999) 257–271.