



UNIVERSIDAD  
DE GRANADA

Escuela Técnica Superior de Ingeniería Informática y de  
Telecomunicación y Facultad de Ciencias

DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y  
MATEMÁTICAS

TRABAJO DE FIN DE GRADO

# Combinando distintas técnicas para el diseño de una metaheurística para problemas de optimización de alta dimensionalidad

Presentado por:  
Jesús García León

**Responsable de  
tutorización**

Daniel Molina Cabrera  
*Departamento de Ciencias de la Computación  
e Inteligencia Artificial*

Curso académico 2023-2024



#### DECLARACIÓN DE ORIGINALIDAD

D./Dña. Jesús García León

Declaro explícitamente que el trabajo presentado como Trabajo de Fin de Grado (TFG), correspondiente al curso académico 2023-2024, es original, entendido esto en el sentido de que no he utilizado para la elaboración del trabajo fuentes sin citarlas debidamente.

En Granada a 2 de julio de 2024

Fdo: Jesús García León



# Agradecimientos

Agradecimientos (opcional, ver archivo preliminares/agradecimiento.tex).



# Índice general

Agradecimientos	III
Resumen	VII
Abstract	IX
Presupuesto	XI
Introducción	XIII
<b>I. Parte Matemática</b>	<b>1</b>
<b>1. Preliminares</b>	<b>3</b>
<b>2. Optimización numérica</b>	<b>5</b>
2.1. Definiciones . . . . .	5
2.2. Teoremas . . . . .	5
2.3. Dificultades(Alta dimensionalidad) . . . . .	5
2.4. ¿Evolución Diferencial? . . . . .	5
<b>3. Agrupamiento de variables</b>	<b>7</b>
3.1. Definiciones . . . . .	7
3.2. Teoremas . . . . .	8
<b>4. Tests estadísticos</b>	<b>13</b>
4.1. Definición . . . . .	13
4.2. Tests . . . . .	13
4.2.1. Test 1 . . . . .	13
4.2.2. Test 2 . . . . .	13
4.2.3. Test 3 . . . . .	13
4.3. Relevancia en el contexto de este TFG . . . . .	13
<b>II. Parte Informática</b>	<b>15</b>
<b>5. Metaheurísticas</b>	<b>17</b>
5.1. Definición . . . . .	17
5.2. Algoritmos evolutivos . . . . .	17
5.2.1. Evolución diferencial . . . . .	17
5.3. Algoritmos de descomposición . . . . .	18
5.4. Búsqueda local . . . . .	18

<b>6. Algoritmos de comparación</b>	<b>19</b>
6.1. SHADE . . . . .	19
6.2. SHADE-ILS . . . . .	23
6.3. DG2 . . . . .	24
6.4. ERDG . . . . .	24
<b>7. Propuesta</b>	<b>25</b>
7.1. DG2-SHADE-ILS . . . . .	25
7.2. ERDG-SHADE-ILS . . . . .	25
<b>8. Resultados</b>	<b>27</b>
8.1. Resultados obtenidos . . . . .	27
<b>9. Conclusiones</b>	<b>29</b>
9.1. Conclusiones extraídas del análisis de los datos . . . . .	29
<b>A. Ejemplo de apéndice</b>	<b>31</b>
<b>Glosario</b>	<b>33</b>
<b>Bibliografía</b>	<b>35</b>



## Resumen

En este trabajo se pretende estudiar las propiedades teóricas de los algoritmos utilizados para resolver problemas de optimización, poniendo un énfasis en los algoritmos utilizados para la optimización en alta dimensión, debido al crecimiento exponencial en complejidad que el aumento de dimensión suele acarrear. También se analizarán distintos test estadísticos y sus propiedades teóricas. Además, se pretende iniciar una biblioteca de algoritmos para dicho tipo de problemas. Se utilizarán técnicas metaheurísticas como algoritmos evolutivos basados en el algoritmo SHADE y se utilizarán técnicas de agrupamiento de variables que permitan descomponer el problema en subproblemas independientes. Combinando estas técnicas, se pretende desarrollar un algoritmo que supere a los anteriores. Finalmente se comparará el algoritmo resultante con sus versiones básicas para comprobar si efectivamente se obtienen mejores resultados.

File: preliminares/resumen.tex



## **Abstract**

An english summary of the project (around 800 and 1500 words are recommended).

File: preliminares/summary.tex



## Presupuesto

Aquí irá el presupuesto estimado del proyecto, basándose en el sueldo de un programador que deberíamos contratar para realizar el proyecto en función del número de horas y el coste de los servidores necesarios para realizar los cálculos de los algoritmos utilizados.

File: preliminares/presupuesto.tex



## Introducción

De acuerdo con la comisión de grado, el TFG debe incluir una introducción en la que se describan claramente los objetivos previstos inicialmente en la propuesta de TFG, indicando si han sido o no alcanzados, los antecedentes importantes para el desarrollo, los resultados obtenidos, en su caso y las principales fuentes consultadas.

Ver archivo preliminares/introduccion.tex





**Parte I.**

**Parte Matemática**



## 1. Preliminares

**Derivada Direccional:** Sea  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función diferenciable, y  $\mathbf{u} = (u_1, \dots, u_n)$  un vector de  $U_X$ . La derivada direccional de  $f$  en la dirección  $\mathbf{u}$ , denotada  $D_{\mathbf{u}}f(\mathbf{x})$ , se define como

$$D_{\mathbf{u}}f(\mathbf{x}) = \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} u_i.$$

**Integral de Línea:** Sea  $L$  una curva con puntos extremos  $A$  y  $B$  en el espacio de decisión  $\mathbb{R}^n$  y la longitud del arco de  $L$  sea  $l$ . Sea  $C$  cualquier punto en  $L$  y la coordenada de  $C(\mathbf{x})$  puede ser determinada de manera única por la longitud del arco  $AC(s)$ :  $\mathbf{x} = \mathbf{x}(s)$ ,  $s \in [0, l]$ . La integral de una función  $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  a lo largo de la curva  $L$  se da por

$$\int_L g(\mathbf{x}) ds = \int_0^l g(\mathbf{x}(s)) ds.$$



## **2. Optimización numérica**

### **2.1. Definiciones**

### **2.2. Teoremas**

### **2.3. Dificultades(Alta dimensionalidad)**

### **2.4. ¿Evolución Diferencial?**



### 3. Agrupamiento de variables

Inspirados por la metodología del análisis clúster, estudiaremos los fundamentos teóricos del agrupamiento diferencial, una técnica de agrupamiento de variables que permite descomponer un problema en subproblemas menores. Esta metodología pretende dividir un conjunto de variables acorde a la interdependencia que se establece entre ellas cuando se trata de optimizar una función objetivo.

Al igual que en el análisis clúster, que agrupa conjuntos de datos en clústers de forma que en cada clúster los datos sean lo más parecidos posibles y distintos del resto de clúster, el objetivo de esta técnica es separar las variables en conjuntos, de forma que cada conjunto sea independiente del resto y dentro de cada conjunto ninguna variable sea independiente. La principal diferencia radica en que en el análisis clúster agrupamos datos acorde al valor de las variables y en el agrupamiento de variables lo que agrupamos son las variables acorde a la dependencia que existe entre ellas.

A continuación, se exponen las definiciones y teoremas necesarios para entender el agrupamiento diferencial desde un punto de vista teórico. En la segunda parte de este TFG, se implementarán distintas variantes de esta técnica para probar su efectividad a la hora de hibridarlas con algoritmos que permitan optimizar una función objetivo.

#### 3.1. Definiciones

**Definición 3.1.** Una función  $f(x_1, \dots, x_n)$  es separable si y solo si:

$$\arg \min_{(x_1, \dots, x_n)} f(x_1, \dots, x_n) = \left( \arg \min_{x_1} f(x_1, \dots), \dots, \arg \min_{x_n} f(\dots, x_n) \right)$$

y no separable en otro caso.

Es decir, podemos encontrar el óptimo de esa función optimizando en cada dimensión por separado.

**Definición 3.2.** Una función  $f(x)$  se dice parcialmente separable con  $m$  componentes si y solo si:

$$\arg \min_x f(x) = \left( \arg \min_{x_1} f(x_1, \dots), \dots, \arg \min_{x_m} f(\dots, x_m) \right)$$

donde  $x = (x_1, \dots, x_n)$  es un vector de decisión de  $n$  dimensiones,  $x_1, \dots, x_n$  son subvectores disjuntos de  $x$  y  $2 \leq m \leq n$ .

La definición 3.2 difiere de 3.1 en que ahora los vectores  $x_i$  no son unidimensionales, es decir, podemos encontrar el óptimo optimizando cada subvector por separado, pero cada subvector puede requerir optimizar un conjunto de variables en vez de una única variable cada vez. La definición 3.1 es un caso particular de 3.2 cuando  $n = m$ .

### 3. Agrupamiento de variables

**Definición 3.3.** Una función es parcialmente aditivamente separable si es de la siguiente forma:

$$f(x) = \sum_{i=1}^m f_i(x_i), \quad m > 1$$

donde  $f_i(\cdot)$  son subfunciones que solo dependen de las variables que forman cada vector  $x_i$  y  $x$  y  $x_i$  se definen como en 3.2. En el caso en el que  $n = m$ , la función se dice totalmente aditivamente separable.

La definición 3.3 es un caso especial de la definición 3.2. Un ejemplo de este tipo de funciones puede ser la función  $f(x) = x_1^2 + x_2^2$ . Es claro que  $f(x) = f_1(x_1) + f_2(x_2)$  con  $f_1(x_1) = x_1^2$  y  $f_2(x_2) = x_2^2$ .

**Definición 3.4.** Se dice que dos variables  $x$  e  $y$  interactúan, si no pueden ser optimizadas de forma independiente. Lo denotaremos por  $x \leftrightarrow y$ .

Introducimos ahora definiciones que serán útiles para detectar variables que interactúan entre sí. La interacción entre variables es otro enfoque de la separabilidad de funciones, pero que nos será útil a la hora de diseñar el algoritmo de agrupamiento diferencial recursivo.

**Definición 3.5.** Sea  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función diferenciable. Las variables de decisión  $x_i$  y  $x_j$  interactúan si existe una solución candidata  $\mathbf{x}^*$  tal que

$$\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i \partial x_j} \neq 0,$$

y diremos que interactúan condicionalmente si

$$\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i \partial x_j} = 0,$$

y existe un conjunto de variables de decisión  $\{x_{k1}, \dots, x_{kt}\} \subset X$ , tal que  $x_i \leftrightarrow x_{k1} \leftrightarrow \dots \leftrightarrow x_{kt} \leftrightarrow x_j$ .

Las variables de decisión  $x_i$  y  $x_j$  son independientes si para cualquier solución candidata  $\mathbf{x}^*$ , se cumple la ecuación anterior y no existe un conjunto de variables de decisión  $\{x_{k1}, \dots, x_{kt}\} \subset X$ , tal que  $x_i \leftrightarrow x_{k1} \leftrightarrow \dots \leftrightarrow x_{kt} \leftrightarrow x_j$ .

## 3.2. Teoremas

**Teorema 3.6.** Sea  $f(\mathbf{x})$  una función (parcialmente) aditivamente separable. Para todo  $a, b_1 \neq b_2, \delta \in \mathbb{R}, \delta \neq 0$ , si se cumple la siguiente condición:

$$\Delta_{\delta, x_p}[f](\mathbf{x})|_{x_p=a, x_q=b_1} \neq \Delta_{\delta, x_p}[f](\mathbf{x})|_{x_p=a, x_q=b_2} \quad (3.1)$$

entonces  $x_p$  y  $x_q$  son no separables, donde

$$\Delta_{\delta, x_p}[f](\mathbf{x}) = f(\dots, x_p + \delta, \dots) - f(\dots, x_p, \dots)$$

se refiere a la diferencia hacia adelante de  $f$  con respecto a la variable  $x_p$  con intervalo  $\delta$ .



El Teorema 3.6 sencillamente nos dice que, dada una función aditivamente separable  $f(\mathbf{x})$ , dos variables  $x_p$  y  $x_q$  interactúan si la diferencia hacia adelante evaluada con dos valores diferentes para  $x_q$  produce resultados diferentes.

Para probar el teorema es suficiente demostrar su contrarrecíproco, que establece que si dos variables  $x_p$  y  $x_q$  son separables, entonces la diferencia hacia adelante evaluada con dos valores diferentes para  $x_q$  produce el mismo resultado.

**Lema 3.7.** Si  $f(\mathbf{x})$  es aditivamente separable, entonces para cualquier  $x_p \in \mathbf{x}$  tenemos

$$\frac{\partial f(\mathbf{x})}{\partial x_p} = \frac{\partial f_i(\mathbf{x}_i)}{\partial x_p}, \quad \forall x_p \in \mathbf{x}_i.$$

*Demostración.* Dado que  $f(\mathbf{x})$  es aditivamente separable, tenemos

$$\frac{\partial f(\mathbf{x})}{\partial x_p} = \frac{\partial}{\partial x_p} \sum_{i=1}^m f_i(\mathbf{x}_i) = \frac{\partial f_1(\mathbf{x}_1)}{\partial x_p} + \dots + \frac{\partial f_m(\mathbf{x}_m)}{\partial x_p}, \quad \forall x_p \in \mathbf{x}_i$$

donde  $\mathbf{x}_1, \dots, \mathbf{x}_m$  son vectores de decisión mutuamente excluyentes. Por lo tanto,

$$\frac{\partial f(\mathbf{x}_j)}{\partial x_p} = 0, \quad \forall j \neq i.$$

Así,

$$\frac{\partial f(\mathbf{x})}{\partial x_p} = \frac{\partial f_i(\mathbf{x}_i)}{\partial x_p}, \quad \forall x_p \in \mathbf{x}_i.$$

□

*Demostración del Teorema 3.6.* Según el Lema 3.7,

$$\frac{\partial f(\mathbf{x})}{\partial x_p} = \frac{\partial f_i(\mathbf{x}_i)}{\partial x_p}, \quad \forall x_p \in \mathbf{x}_i.$$

Entonces, para todo  $x_q \notin \mathbf{x}_i$  tenemos

$$\left. \frac{\partial f(\mathbf{x})}{\partial x_p} \right|_{x_q=b_1} = \left. \frac{\partial f(\mathbf{x})}{\partial x_p} \right|_{x_q=b_2} = \frac{\partial f_i(\mathbf{x}_i)}{\partial x_p}, \quad \forall b_1 \neq b_2.$$

$$\int_a^{a+\delta} \left. \frac{\partial f(\mathbf{x})}{\partial x_p} dx_p \right|_{x_q=b_1} = \int_a^{a+\delta} \left. \frac{\partial f(\mathbf{x})}{\partial x_p} dx_p \right|_{x_q=b_2}$$

$$\Delta_{\delta, x_p}[f](\mathbf{x})|_{x_p=a, x_q=b_1} = \Delta_{\delta, x_p}[f](\mathbf{x})|_{x_p=a, x_q=b_2} \quad \forall a, b_1 \neq b_2, \delta \in \mathbb{R}, \delta \neq 0.$$

□

El Teorema 3.6 es el que nos servirá en la segunda parte para diseñar el algoritmo de agrupamiento diferencial.

**Notación:** Sea  $X$  el conjunto de variables de decisión  $\{x_1, \dots, x_n\}$  y  $U_X$  el conjunto de vectores unitarios en el espacio de decisión  $\mathbb{R}^n$ . Sea  $X_1$  un subconjunto de variables de decisión  $X_1 \subset X$  y  $U_{X_1}$  un subconjunto de  $U_X$  tal que para cualquier vector unitario  $\mathbf{u} = (u_1, \dots, u_n) \in U_{X_1}$ , tenemos  $u_i = 0$  si  $x_i \notin X_1$ .

### 3. Agrupamiento de variables

**Proposición 3.8.** Sea  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función diferenciable;  $X_1 \subset X$  y  $X_2 \subset X$  dos subconjuntos mutuamente excluyentes de variables de decisión:  $X_1 \cap X_2 = \emptyset$ . Si existen dos vectores unitarios  $\mathbf{u}_1 \in U_{X_1}$  y  $\mathbf{u}_2 \in U_{X_2}$ , y una solución candidata  $\mathbf{x}^*$  en el espacio de decisión tal que

$$D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}^*) \neq 0$$

entonces hay alguna interacción entre las variables de decisión en  $X_1$  y  $X_2$ .

*Demostración.* Sin pérdida de generalidad, asumimos que  $X_1 = \{x_{1,1}, \dots, x_{1,p}\}$ ,  $X_2 = \{x_{2,1}, \dots, x_{2,q}\}$ , donde  $p$  y  $q$  son el número de variables de decisión en  $X_1$  y  $X_2$ , respectivamente;  $\mathbf{u}_1 = (u_1^1, \dots, u_1^n)$  y  $\mathbf{u}_2 = (u_2^1, \dots, u_2^n)$ . Según la derivada direccional,

$$D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} u_1^i u_2^j.$$

Como  $\mathbf{u}_1$  y  $\mathbf{u}_2$  son dos vectores unitarios de  $U_{X_1}$  y  $U_{X_2}$ , respectivamente, podemos obtener que

$$\begin{aligned} u_1^i &= 0, \text{ si } x_i \notin X_1, \\ u_2^j &= 0, \text{ si } x_j \notin X_2. \end{aligned}$$

Por lo tanto,

$$D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}) = \sum_{i=1}^p \sum_{j=1}^q \frac{\partial^2 f(\mathbf{x})}{\partial x_{1,i} \partial x_{2,j}} u_1^{1,i} u_2^{2,j}.$$

Si se cumple (3.1),

$$\sum_{i=1}^p \sum_{j=1}^q \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_{1,i} \partial x_{2,j}} u_1^{1,i} u_2^{2,j} \neq 0.$$

Por lo tanto, existe al menos un par  $(i, j)$ , tal que

$$\frac{\partial^2 f(\mathbf{x}^*)}{\partial x_{1,i} \partial x_{2,j}} \neq 0.$$

Basado en la Definición 3.5, al menos un par de variables de decisión  $x_{1,i} \in X_1$  y  $x_{2,j} \in X_2$  interactúan.  $\square$

**Corolario 3.9.** Sea  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  una función objetivo;  $X_1 \subset X$  y  $X_2 \subset X$  dos subconjuntos mutuamente excluyentes de variables de decisión:  $X_1 \cap X_2 = \emptyset$ . Si existen dos vectores unitarios  $\mathbf{u}_1 \in U_{X_1}$  y  $\mathbf{u}_2 \in U_{X_2}$ , dos números reales  $l_1, l_2 > 0$ , y una solución candidata  $\mathbf{x}^*$  en el espacio de decisión, tal que

$$f(\mathbf{x}^* + l_1 \mathbf{u}_1 + l_2 \mathbf{u}_2) - f(\mathbf{x}^* + l_2 \mathbf{u}_2) \neq f(\mathbf{x}^* + l_1 \mathbf{u}_1) - f(\mathbf{x}^*) \quad (3.2)$$

entonces hay alguna interacción entre las variables de decisión en  $X_1$  y  $X_2$ .

*Demostración.* Con la Proposición 3.8, solo necesitamos probar la siguiente afirmación.

**Afirmación 1:** Si existen dos vectores unitarios  $\mathbf{u}_1 \in U_{X_1}$  y  $\mathbf{u}_2 \in U_{X_2}$ , dos números reales  $l_1, l_2 > 0$ , y una solución candidata  $\mathbf{x}^*$  en el espacio de decisión, tal que (3.2) se cumple, entonces (3.1) es verdadero.

Es equivalente probar su contrarecíproco.

**Afirmación 2:** Si para cualquier par de vectores unitarios  $\mathbf{u}_1 \in U_{X_1}$  y  $\mathbf{u}_2 \in U_{X_2}$ , y para cualquier solución candidata  $\mathbf{x}^*$  en el espacio de decisión, se cumple la siguiente condición:

$$D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}^*) = 0 \quad (3.3)$$

entonces

$$f(\mathbf{x}^* + l_1 \mathbf{u}_1 + l_2 \mathbf{u}_2) - f(\mathbf{x}^* + l_2 \mathbf{u}_2) = f(\mathbf{x}^* + l_1 \mathbf{u}_1) - f(\mathbf{x}^*)$$

para cualquier  $l_1, l_2 > 0$ .

Sea  $A_2(\mathbf{x}^*)$  cualquier punto en  $\mathbb{R}^n$ , y  $B_2$  sea  $\mathbf{x}^* + l_2 \mathbf{u}_2$ , donde  $\mathbf{u}_2$  es cualquier vector en  $U_{X_2}$  y  $l_2$  es cualquier número real positivo. Sea  $C_2$  cualquier punto en el segmento  $A_2 B_2$ . Por lo tanto, la longitud del segmento  $A_2 B_2$  es  $l_2$ , y la coordenada de  $C_2(\mathbf{x})$  puede ser determinada de manera única por la longitud del segmento  $A_2 C_2(s_2)$ :  $\mathbf{x}(s_2) = \mathbf{x}^* + s_2 \mathbf{u}_2$ ,  $s_2 \in [0, l_2]$ . Si (3.3) se cumple para cualquier solución candidata en el espacio de decisión, entonces

$$D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}) = 0. \quad (3.4)$$

Como  $D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}) = D_{\mathbf{u}_2} D_{\mathbf{u}_1} f(\mathbf{x})$ , integrando ambos lados de (3.4) a lo largo del segmento  $A_2 B_2$ , obtenemos

$$\int_0^{l_2} D_{\mathbf{u}_1} D_{\mathbf{u}_2} f(\mathbf{x}) ds_2 = \int_0^{l_2} D_{\mathbf{u}_2} D_{\mathbf{u}_1} f(\mathbf{x}) ds_2 = 0.$$

Como

$$\int_0^{l_2} D_{\mathbf{u}_2} (D_{\mathbf{u}_1} f(\mathbf{x}(s_2))) ds_2 = D_{\mathbf{u}_1} f(\mathbf{x}(s_2)) \Big|_{s_2=0}^{s_2=l_2},$$

entonces,

$$D_{\mathbf{u}_1} f(\mathbf{x}(s_2)) \Big|_{s_2=0}^{s_2=l_2} = 0$$

y

$$D_{\mathbf{u}_1} f(\mathbf{x}^* + l_2 \mathbf{u}_2) - D_{\mathbf{u}_1} f(\mathbf{x}^*) = 0.$$

Como  $A_2(\mathbf{x}^*)$  es cualquier punto en  $\mathbb{R}^n$ , entonces

$$D_{\mathbf{u}_1} f(\mathbf{x} + l_2 \mathbf{u}_2) = D_{\mathbf{u}_1} f(\mathbf{x}). \quad (3.5)$$

Sea  $A_1(\mathbf{x}^*)$  cualquier punto en  $\mathbb{R}^n$ , y  $B_1$  sea  $\mathbf{x}^* + l_1 \mathbf{u}_1$ , donde  $\mathbf{u}_1$  es cualquier vector en  $U_{X_1}$  y  $l_1$  es cualquier número real positivo. Sea  $C_1$  cualquier punto en el segmento  $A_1 B_1$ . Por lo tanto, la longitud del segmento  $A_1 B_1$  es  $l_1$ , y la coordenada de  $C_1(\mathbf{x})$  puede ser determinada de manera única por la longitud del segmento  $A_1 C_1(s_1)$ :  $\mathbf{x}(s_1) = \mathbf{x}^* + s_1 \mathbf{u}_1$ ,  $s_1 \in [0, l_1]$ . De manera similar, integrando ambos lados de (3.5) a lo largo del segmento  $A_1 B_1$ , obtenemos

$$\int_0^{l_1} D_{\mathbf{u}_1} f(\mathbf{x}(s_1) + l_2 \mathbf{u}_2) ds_1 = \int_0^{l_1} D_{\mathbf{u}_1} f(\mathbf{x}(s_1)) ds_1.$$

Por lo tanto,

$$f(\mathbf{x}^* + l_1 \mathbf{u}_1 + l_2 \mathbf{u}_2) - f(\mathbf{x}^* + l_2 \mathbf{u}_2) = f(\mathbf{x}^* + l_1 \mathbf{u}_1) - f(\mathbf{x}^*).$$

Así, la Afirmación 2 queda probada, y la Afirmación 1 y el Corolario 3.9 son verdaderos.  $\square$



## **4. Tests estadísticos**

### **4.1. Definición**

### **4.2. Tests**

#### **4.2.1. Test 1**

#### **4.2.2. Test 2**

#### **4.2.3. Test 3**

### **4.3. Relevancia en el contexto de este TFG**



**Parte II.**

**Parte Informática**





## 5. Metaheurísticas

### 5.1. Definición

Explicar las metaheurísticas en general, diferencias con las heurísticas y su clasificación. Clasificación general de las metaheurísticas para LSGO y centrarnos en algoritmos evolutivos, meméticos y evolución diferencial agrupamiento de variables

### 5.2. Algoritmos evolutivos

#### 5.2.1. Evolución diferencial

Esta sección describe brevemente la Evolución Diferencial (DE). Similar a otros algoritmos evolutivos para la optimización numérica, una población de DE se representa como un conjunto de vectores de parámetros reales  $\mathbf{x}_i = (x_1, \dots, x_D), i = 1, \dots, N$ , donde  $D$  es la dimensionalidad del problema objetivo y  $N$  es el tamaño de la población.

Al comienzo de la búsqueda, los vectores individuales de la población se inicializan aleatoriamente. Luego, se repite un proceso de generación de vectores de prueba y selección hasta que se encuentra algún criterio de parada. En cada generación  $G$ , se genera un vector mutado  $\mathbf{v}_{i,G}$  a partir de un miembro de la población existente  $\mathbf{x}_{i,G}$  aplicando alguna estrategia de mutación. A continuación se muestran ejemplos de estrategias de mutación:

- **rand/1**

$$\mathbf{v}_{i,G} = \mathbf{x}_{r1,G} + F \cdot (\mathbf{x}_{r2,G} - \mathbf{x}_{r3,G})$$

- **rand/2**

$$\mathbf{v}_{i,G} = \mathbf{x}_{r1,G} + F \cdot (\mathbf{x}_{r2,G} - \mathbf{x}_{r3,G}) + F \cdot (\mathbf{x}_{r4,G} - \mathbf{x}_{r5,G})$$

- **best/1**

$$\mathbf{v}_{i,G} = \mathbf{x}_{best,G} + F \cdot (\mathbf{x}_{r1,G} - \mathbf{x}_{r2,G})$$

- **current-to-best/1**

$$\mathbf{v}_{i,G} = \mathbf{x}_{i,G} + F \cdot (\mathbf{x}_{best,G} - \mathbf{x}_{i,G}) + F \cdot (\mathbf{x}_{r1,G} - \mathbf{x}_{r2,G})$$

Los índices  $r1, \dots, r5$  se seleccionan aleatoriamente de  $[1, N]$  de manera que difieran entre sí, así como  $i$ .  $\mathbf{x}_{best,G}$  es el mejor individuo en la población en la generación  $G$ . El parámetro  $F \in [0, 1]$  controla la potencia del operador de mutación diferencial, a mayor  $F$ , mayor será la diferencia entre el vector original y el mutado.

Después de generar el vector mutado  $\mathbf{v}_{i,G}$ , se cruza con el padre  $\mathbf{x}_{i,G}$  para generar el vector de prueba  $\mathbf{u}_{i,G}$ . El cruce binomial, el operador de cruce más utilizado en DE, se implementa de la siguiente manera:

$$u_{j,i,G} = \begin{cases} v_{j,i,G} & \text{si } \text{rand}[0, 1) \leq CR \text{ o } j = j_{\text{rand}} \\ x_{j,i,G} & \text{de lo contrario} \end{cases}$$

$\text{rand}[0, 1)$  denota un número aleatorio seleccionado uniformemente de  $[0, 1)$ , y  $j_{\text{rand}}$  es un índice de variable de decisión que se selecciona aleatoriamente y de manera uniforme de  $[1, D]$ .  $CR \in [0, 1]$  es la tasa de cruce.

Después de que se han generado todos los vectores de prueba  $\mathbf{u}_{i,G}$ , un proceso de selección determina los supervivientes para la siguiente generación. El operador de selección en DE estándar compara cada individuo  $\mathbf{x}_{i,G}$  contra su vector de prueba correspondiente  $\mathbf{u}_{i,G}$ , manteniendo el mejor vector en la población.

$$\mathbf{x}_{i,G+1} = \begin{cases} \mathbf{u}_{i,G} & \text{si } f(\mathbf{u}_{i,G}) \leq f(\mathbf{x}_{i,G}) \\ \mathbf{x}_{i,G} & \text{de lo contrario} \end{cases}$$

### 5.3. Algoritmos de descomposición

### 5.4. Búsqueda local

## 6. Algoritmos de comparación

Se explican los algoritmos que combinaremos para crear nuestra propuesta y que se utilizarán también para comparar como mejora el algoritmo final comparado con los algoritmos básicos. Se incluirá el pseudocódigo y la explicación de las partes esenciales que componen cada algoritmo.

### 6.1. SHADE

En esta sección explicaremos el algoritmo SHADE (Success-history based parameter adaptation for Differential Evolution), que es un componente clave del algoritmo SHADE-ILS que utilizaremos en nuestra propuesta. El artículo original donde se publicó este algoritmo puede encontrarse en [TF13].

SHADE (Success-History based Adaptive Differential Evolution) es una variante del algoritmo de Evolución Diferencial (DE) que utiliza un esquema de adaptación de parámetros basado en el historial de éxitos. A diferencia de otras variantes de DE, SHADE mantiene una memoria histórica de los valores de los parámetros de control que han sido exitosos en generaciones anteriores, y utiliza esta información para guiar la selección de los parámetros de control en generaciones futuras. El objetivo es mejorar la eficiencia de búsqueda y la capacidad de encontrar soluciones óptimas en problemas de optimización. Sus componentes principales que lo diferencian de la evolución estándar se describen a continuación.

#### Estrategia de mutación

La estrategia de mutación utilizada por SHADE es denominada current-to-p-best/1.

- Estrategia current-to-pbest/1:

$$v_{i,G} = x_{i,G} + F_i \cdot (x_{\text{pbest},G} - x_{i,G}) + F_i \cdot (x_{r1,G} - x_{r2,G})$$

El individuo  $x_{\text{pbest},G}$  es seleccionado del  $N \cdot p$  ( $p \in [0, 1]$ ) mejor de la generación  $G$ .  $F_i$  es el parámetro  $F$  usado por el individuo  $x_i$ . Este parámetro  $p$  controla la voracidad del algoritmo, para balancear exploración con explotación. A menor  $p$ , mayor explotación.

En SHADE, cada individuo  $x_i$  tiene un  $p_i$  asociado, que se establece según la siguiente ecuación por generación:

$$p_i = \text{rand}[p_{\min}, 0.2]$$

donde  $p_{\min}$  se establece de manera que cuando se selecciona el mejor individuo pbest, se seleccionen al menos 2 individuos, es decir,  $p_{\min} = 2/N$ . El valor máximo de 0.2 en la ecuación 6.1 es el valor máximo del rango para  $p$  sugerido.

### Archivo Externo

Para mantener la diversidad, SHADE utiliza un archivo externo opcional. Los vectores padres  $x_{i,G}$  que fueron peores que los vectores de prueba  $u_{i,G}$  (y por lo tanto no son seleccionados para la supervivencia en el DE estándar) son preservados. Cuando se usa el archivo,  $x_{r2,G}$  en la ecuación de mutación 6.1 es seleccionado de  $P \cup A$ , la unión de la población  $P$  y el archivo  $A$ . El tamaño del archivo se establece igual al de la población, es decir,  $|A| = |P|$ . Siempre que el tamaño del archivo excede  $|A|$ , los elementos seleccionados aleatoriamente son eliminados para hacer espacio para los nuevos elementos insertados.

### Adaptación de Parámetros

SHADE utiliza un mecanismo de adaptación de parámetros basado en un registro histórico de configuraciones de parámetros exitosas. Para ello, SHADE mantiene una memoria histórica con  $H$  entradas para ambos parámetros de control de DE,  $CR$  y  $F$ ,  $M_{CR}$  y  $M_F$ . Una representación de esta tabla se puede ver en 6.1. Al comienzo, el contenido de  $M_{CR,i}$  y  $M_{F,i}$  ( $i = 1, \dots, H$ ) se inicializa a 0.5.

En cada generación, los parámetros de control  $CR_i$  y  $F_i$  utilizados por cada individuo  $x_i$  se generan seleccionando primero un índice  $r_i$  aleatoriamente de  $[1, H]$ , y luego aplicando las siguientes ecuaciones:

$$CR_i = \text{randn}(M_{CR,r_i}, 0.1)$$

$$F_i = \text{randc}(M_{F,r_i}, 0.1)$$

Aquí,  $\text{randn}(\mu, \sigma^2)$  y  $\text{randc}(\mu, \sigma^2)$  son distribuciones normales y de Cauchy, respectivamente, con media  $\mu$  y varianza  $\sigma^2$ .

Si los valores generados para  $CR_i$  están fuera del rango  $[0, 1]$ , se reemplazan por el valor límite (0 o 1) más cercano al valor generado. Cuando  $F_i > 1$ ,  $F_i$  se trunca a 1, y cuando  $F_i \leq 0$ , se aplica repetidamente la ecuación 6.1 para intentar generar un valor válido.

En cada generación, los valores  $CR_i$  y  $F_i$  que logran generar un vector de prueba  $u_{i,G}$  que es mejor que el individuo padre  $x_{i,G}$  se registran como  $S_{CR}$  y  $S_F$ .

Los valores medios de  $S_{CR}$  y  $S_F$  para cada generación se almacenan en una memoria histórica  $M_{CR}$  y  $M_F$ . SHADE mantiene un conjunto diverso de parámetros para guiar la adaptación de parámetros de control a medida que avanza la búsqueda. Por lo tanto, incluso si  $S_{CR}$  y  $S_F$  para alguna generación particular contienen un conjunto deficiente de valores, los parámetros almacenados en la memoria de generaciones anteriores no pueden verse directamente afectados de manera negativa.

Al final de la generación, el contenido de la memoria se actualiza de la siguiente manera:

$$M_{CR,k,G+1} = \begin{cases} \text{meanWA}(S_{CR}) & \text{si } S_{CR} \neq \emptyset \\ M_{CR,k,G} & \text{de lo contrario} \end{cases}$$

$$M_{F,k,G+1} = \begin{cases} \text{meanWL}(S_F) & \text{si } S_F \neq \emptyset \\ M_{F,k,G} & \text{de lo contrario} \end{cases}$$

Un índice  $k$  ( $1 \leq k \leq H$ ) determina la posición en la memoria a actualizar. Al comienzo de la búsqueda,  $k$  se inicializa a 1.  $k$  se incrementa cada vez que se inserta un nuevo elemento en el historial. Si  $k > H$ ,  $k$  se establece en 1. En la generación  $G$ , se actualiza el  $k$ -ésimo elemento en la memoria. Nótese que cuando todos los individuos en la generación  $G$  no logran generar

un vector de prueba que sea mejor que el padre, es decir,  $S_{CR} = S_F = \emptyset$ , la memoria no se actualiza.

Además, la media ponderada  $\text{meanWA}(S_{CR})$  y la media ponderada de Lehmer  $\text{meanWL}(S_F)$  se calculan usando las fórmulas descritas a continuación, y al igual que  $\text{meanWA}(S_{CR})$ , la cantidad de mejora se usa para influir en la adaptación de parámetros.

$$\text{meanWA}(S_{CR}) = \sum_{k=1}^{|S_{CR}|} w_k \cdot S_{CR,k}$$

$$w_k = \frac{\Delta f_k}{\sum_{k=1}^{|S_{CR}|} \Delta f_k}$$

donde  $\Delta f_k = |f(u_{k,G}) - f(x_{k,G})|$ .

$$\text{meanWL}(S_F) = \frac{\sum_{k=1}^{|S_F|} w_k \cdot S_{F,k}^2}{\sum_{k=1}^{|S_F|} w_k \cdot S_{F,k}}$$

Index	1	2	...	H - 1	H
$M_{CR}$	$M_{CR,1}$	$M_{CR,2}$	...	$M_{CR,H-1}$	$M_{CR,H}$
$M_F$	$M_{F,1}$	$M_{F,2}$	...	$M_{F,H-1}$	$M_{F,H}$

Tabla 6.1.: La memoria histórica  $M_{CR}$ ,  $M_F$

**Pseudocódigo de SHADE****Algorithm 1** SHADE

---

```

1: // Fase de inicialización
2:  $G = 0$ ;
3: Inicializar población  $P_0 = (x_{1,0}, \dots, x_{N,0})$  aleatoriamente;
4: Establecer todos los valores en  $M_{CR}$ ,  $M_F$  a 0.5;
5: Archivo  $A = \emptyset$ ;
6: Contador de índice  $k = 1$ ;
7: // Bucle principal
8: while No se cumplen los criterios de terminación do
9:    $S_{CR} = \emptyset$ ;  $S_F = \emptyset$ ;
10:  for  $i = 1$  to  $N$  do
11:     $r_i =$  Seleccionar aleatoriamente de  $[1, H]$ ;
12:     $CR_{i,G} = \text{randn}_i(M_{CR,r_i}, 0.1)$ ;
13:     $F_{i,G} = \text{randc}_i(M_F, r_i, 0.1)$ ;
14:     $p_{i,G} = \text{rand}[p_{\min}, 0.2]$ ;
15:    Generar vector de prueba  $u_{i,G}$  usando current-to-pbest/1/bin;
16:  end for
17:  for  $i = 1$  to  $N$  do
18:    if  $f(u_{i,G}) \leq f(x_{i,G})$  then
19:       $x_{i,G+1} = u_{i,G}$ ;
20:    else
21:       $x_{i,G+1} = x_{i,G}$ ;
22:    end if
23:    if  $f(u_{i,G}) < f(x_{i,G})$  then
24:       $x_{i,G} \rightarrow A$ ;
25:       $CR_{i,G} \rightarrow S_{CR}$ ,  $F_{i,G} \rightarrow S_F$ ;
26:    end if
27:  end for
28:  Si  $|A| \geq |P|$ , se eliminan individuos seleccionados aleatoriamente para que  $|A| \leq |P|$ ;
29:  if  $S_{CR} \neq \emptyset$  y  $S_F \neq \emptyset$  then
30:    Actualizar  $M_{CR,k}$ ,  $M_F,k$  basado en  $S_{CR}$ ,  $S_F$ ;
31:     $k = k + 1$ ;
32:    if  $k > H$  then
33:       $k$  se establece en 1;
34:    end if
35:  end if
36: end while

```

---

## 6.2. SHADE-ILS

En este apartado, presentamos el algoritmo SHADE-ILS, expuesto por primera vez en [MLH18], que combina el uso de la técnica de evolución diferencial SHADE explicada anteriormente y la búsqueda local. Además proporciona un método de reinicio para cuando se considera que la población se ha estancado en un óptimo local y no se puede mejorar más. Describiremos en detalle los elementos que componen el algoritmo y un pseudocódigo que nos proporcione una visión global del algoritmo.

---

### Algorithm 2 SHADE-ILS

---

```

1: Algoritmo 1: SHADE-ILS
2: población  $\leftarrow$  random(dim, tamaño_población)
3: solución_inicial  $\leftarrow$  (superior + inferior)/2
4: actual_mejor  $\leftarrow$  BL(solución_inicial)
5: mejor_solución  $\leftarrow$  actual_mejor
6: while evaluaciones_totales < evaluaciones_máximas do
7:   previo  $\leftarrow$  actual_mejor.fitness
8:   actual_mejor  $\leftarrow$  SHADE(población, actual_mejor)
9:   mejora  $\leftarrow$  previo - actual_mejor.fitness
10:  Escoge el método de BL a aplicar en esta iteración.
11:  actual_mejor  $\leftarrow$  BL(población, actual_mejor)
12:  Actualiza probabilidad de aplicar BL.
13:  if mejor(actual_mejor, mejor_solución) then
14:    mejor_solución  $\leftarrow$  actual_mejor
15:  end if
16:  if Debe reiniciar then
17:    Reinicia y actualiza actual_mejor
18:  end if
19: end while

```

---

### Método de exploración

Como su nombre indica, el algoritmo SHADE-ILS utiliza como método de exploración del espacio el algoritmo SHADE, explicado en detalle en la sección anterior.

### Selección de la Búsqueda Local

La selección de la búsqueda local a utilizar en cada iteración se lleva a cabo según la mejora que ha aportado cada búsqueda local en su última aplicación. Inicialmente la mejora de cada método de búsqueda local es 0, así que en las primeras llamadas a la función de búsqueda local se aplique cada vez un método distinto hasta haber aplicado todos una vez. Una vez se ha aplicado un método cada vez, tenemos para cada método el ratio de mejora  $I_{LS}$ . A partir de ahora se aplicará siempre el método con mayor  $I_{LS}$  y se actualizará este valor en cada aplicación de la BL seleccionada. De esta forma se intentará aplicar siempre el método que mayor mejora aporta, cuando un método tenga un rendimiento peor, su  $I_{LS}$  disminuirá y otro método con mayor  $I_{LS}$  ocupará su lugar. Este método no garantiza aplicar siempre el método óptimo, pero proporciona una buena heurística para decidir que método aplicar y permite cambiar rápidamente de método de BL si otro método se estanca. El  $I_{LS}$  se calcula como:

## 6. Algoritmos de comparación

$$ILS = \frac{\text{fitness(BeforeLS)} - \text{fitness(AfterLS)}}{\text{fitness(BeforeLS)}}$$

En [MLH18] se propone utilizar dos métodos de búsqueda local: el algoritmo MTS LS-1 y L-BFGS-B. El primero está especialmente diseñado para problemas LSGO y es apropiado para problemas separables pero es muy sensible a rotaciones, el segundo es menos potente, pero menos sensible a rotaciones.

### Mecanismo de reinicio

El mecanismo de reinicio que se propone en [MLH18] consiste en reiniciar la población cuando se da la condición de que durante tres iteraciones consecutivas, el ratio de mejora es menor del 5 %. En estos casos el mecanismo de reinicio aplicado sigue los siguientes pasos:

- Se selecciona aleatoriamente una solución sol.
- Se aplica una perturbación a sol que siga una distribución uniforme de media 0 y longitud del intervalo un 1 % del dominio de búsqueda:

$$\text{currentbest} = \text{sol} + \text{rand}_i \cdot 0.01 \cdot (b - a)$$

donde  $\text{rand}_i$  devuelve un número aleatorio  $\text{rand}_i \in [-1, 1]$  y  $[a, b]$  es el dominio de búsqueda.

- Los parámetros adaptativos de los métodos de BL se reinician a sus valores por defecto.

## 6.3. DG2

[OYM<sup>+</sup>17]

## 6.4. ERDG



## **7. Propuesta**

### **7.1. DG2-SHADE-ILS**

### **7.2. ERDG-SHADE-ILS**



## **8. Resultados**

### **8.1. Resultados obtenidos**



## **9. Conclusiones**

### **9.1. Conclusiones extraídas del análisis de los datos**



## A. Ejemplo de apéndice

Los apéndices son opcionales.

Este fichero `apendice-ejemplo.tex` es una plantilla para añadir apéndices al TFG. Para ello, es necesario:

- Crear una copia de este fichero `apendice-ejemplo.tex` en la carpeta `apendices` con un nombre apropiado (p.e. `apendice01.tex`).
- Añadir el comando `\input{apendices/apendice01}` en el fichero principal `tfg.tex` donde queremos que aparezca dicho apéndice (debe de ser después del comando `\appendix`).





## Glosario

La inclusión de un glosario es opcional.

Archivo: `glosario.tex`

$\mathbb{R}$  Conjunto de números reales.

$\mathbb{C}$  Conjunto de números complejos.

$\mathbb{Z}$  Conjunto de números enteros.



## Bibliografía

- [MLH18] Daniel Molina, Antonio LaTorre, y Francisco Herrera. Shade with iterative local search for large-scale global optimization. En *Proceedings of the 2018 IEEE Congress on Evolutionary Computation*, páginas 1252–1259, Rio de Janeiro, Brasil, July 2018.
- [OYM<sup>+</sup>17] Mohammad Nabi Omidvar, Ming Yang, Yi Mei, Xiaodong Li, y Xin Yao. Dg2: A faster and more accurate differential grouping for large-scale black-box optimization. *IEEE Transactions on Evolutionary Computation*, 21(6):929–942, 2017.
- [TF13] Ryoji Tanabe y Alex Fukunaga. Success-history based parameter adaptation for differential evolution. En *2013 IEEE Congress on Evolutionary Computation*, páginas 71–78, 2013.