



ugr

Universidad  
de Granada

TRABAJO FIN DE GRADO

DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y MATEMÁTICAS

# Diseño de Metaheurística para problemas combinatorios costosos

---

**Autor**

Irene Trigueros Lorca

**Directores**

Daniel Molina Cabrera

Francisco Herrera Triguero



ESCUELA TÉCNICA SUPERIOR DE  
INGENIERÍAS INFORMÁTICA  
Y DE TELECOMUNICACIÓN



Facultad de Ciencias

FACULTAD DE CIENCIAS

Granada, Mayo de 2023









# Diseño de Metaheurística para problemas combinatorios costosos

---

Subtítulo del proyecto.

## **Autor**

Irene Trigueros Lorca

## **Directores**

Daniel Molina Cabrera

Francisco Herrera Triguero



# Diseño de Metaheurística para problemas combinatorios costosos

Irene Trigueros Lorca

**Palabras clave:** palabra\_clave1, palabra\_clave2, palabra\_clave3, .....

## Resumen

Poner aquí el resumen.





# Metaheuristics design for expensive combinatorial problems

Irene Trigueros Lorca

**Keywords:** Keyword1, Keyword2, Keyword3, ....

## Abstract

Write here the abstract in English.



---

Yo, **Irene Trigueros Lorca**, alumno de la titulación Doble Grado en Ingeniería Informática y Matemáticas de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada** y de la **Facultad de Ciencias**, con DNI 77385991F, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca de ambos centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Irene Trigueros Lorca

Granada a, 3 de mayo de 2023.



---

D. **Daniel Molina Cabrera**, Profesor del Departamento Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

D. **Francisco Herrera Triguero**, Profesor del Departamento Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

**Informan:**

Que el presente trabajo, titulado ***Diseño de Metaheurística para problemas combinatorios costosos***, ha sido realizado bajo su supervisión por **Irene Trigueros Lorca**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 3 de mayo de 2023.

**Los directores:**

**Daniel Molina Cabrera**

**Francisco Herrera Triguero**



# Agradecimientos

Poner aquí agradecimientos...





# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Estimación y Presupuesto</b>	<b>3</b>
2.1. Planificación . . . . .	3
2.1.1. Planificación Base . . . . .	3
2.1.2. Planificación Final . . . . .	4
2.2. Presupuesto . . . . .	5
<b>3. Repaso bibliográfico</b>	<b>7</b>
<b>4. Contexto matemático</b>	<b>9</b>
4.1. Definición del problema . . . . .	9
4.2. Condición de Optimalidad . . . . .	11
4.3. Rendimiento de los algoritmos . . . . .	13
4.3.1. Convergencia y Orden de Convergencia . . . . .	13
4.3.2. Comportamiento numérico . . . . .	14
4.4. Programación no lineal sin restricciones . . . . .	15
4.4.1. Algoritmos de Optimización sin restricciones . . . . .	15
4.4.2. Algoritmos de Búsqueda Lineal . . . . .	16
4.4.3. Métodos de Gradiente . . . . .	17
4.4.4. Métodos de Gradiente Conjugado . . . . .	18
4.4.5. Métodos de Newton . . . . .	19
4.5. Algoritmos de Referencia experimentación . . . . .	35
4.6. Resultados versión expensive . . . . .	35
4.7. Incorporación del histórico . . . . .	35
4.8. Uso de GRASP . . . . .	35
4.9. Operador de Cruce Intensivo . . . . .	35
4.10. Estudio de la diversidad . . . . .	35

4.11. Incrementando la diversidad (con nuevo reemplazo) . . . . .	35
---	----

## Índice de figuras



# Capítulo 1: Introducción



## Capítulo 2: Estimación y Presupuesto

En este capítulo se detalla cómo se ha organizado el trabajo y el tiempo dedicado a cada una de ellas. El orden en el que se han realizado dichas tareas, queda representado en un Diagrama de Gantt () para su mejor comprensión. Además, se realiza una estimación del presupuesto necesario para desarrollar el proyecto.

### 2.1. Planificación

La planificación previa de este proyecto se ha realizado siguiendo una metodología ágil. Es decir, la planificación se va adaptando dependiendo cómo hayan transcurrido las tareas anteriores. Es el modelo de planificación que más se ajusta a este tipo de trabajo, ya que, *a priori*, se desconoce la dificultad de las tareas a realizar.

Sin embargo, es cierto que antes de empezar el trabajo se estableció una planificación base bastante amplia para poder asegurar que se iba a finalizar el proyecto a tiempo.

#### 2.1.1. Planificación Base

La planificación inicial que se estableció antes de iniciar el proyecto se puede expresar mediante la información representada en la tabla 2.1:

Resumen	Pasos	Duración
Investigar el problema	Obtener instancias del problema Buscar algoritmos que lo resuelvan Buscar posibles implementaciones	Noviembre
Algoritmos de referencia	Elegir un algoritmo como referencia y estudiarlo Reducir el problema y meter equilibrio	Enero-Febrero
Experimentación	Formas de inicialización no aleatorias → Diseño experimental	2 semanas
Histórico	Usar el histórico	Marzo
Memoria	Escribir el informe	Mayo

Cuadro 2.1: Planificación Base

### 2.1.2. Planificación Final

#### Tareas realizadas

Si bien es cierto que se han realizado una gran cantidad de tareas (sobre todo distintas modificaciones sobre algoritmos), con el fin de mantener la simplicidad, se han agrupado algunas tareas que tenían funciones similares. Esta agrupación también es de ayuda para simplificar la estimación del tiempo que se le ha dedicado a cada una de las tareas. Así, las tareas realizadas se resumen en la siguiente lista:

1. **Planteamiento y comprensión del problema:** Revisión del trabajo a realizar y reuniones con los tutores para proponer modificaciones y comprender mejor y aclarar todos los matices del Trabajo de Fin de Grado.
2. **Búsqueda de información y lecturas:** Búsqueda y lectura comprensiva de todos los artículos y documentos necesarios para la realización del proyecto.
3. **Planificación del proyecto:** Planificación de algunos aspectos que usar como base, así como las tareas que eran necesarias inicialmente. También hace referencia a partes de reuniones con los tutores para modificar las planificaciones (añadiendo o eliminando tareas) dependiendo del progreso alcanzado y los resultados obtenidos.
4. **Implementación de la propuesta inicial:** Implementación del código de los algoritmos base.
5. **Adaptación de los algoritmos base:** Modificación del código de los algoritmos base para adaptarlos al problema en cuestión.
6. **Modificación de los algoritmos:** Sucesivas modificaciones sobre el algoritmo base y los algoritmos que mejores resultados proporcionaban con el fin de mejorarlos aún más.
7. **Obtención de resultados:** Ejecución del código para obtener todos los resultados y cambiarlos de formato para su posterior análisis.
8. **Análisis de los resultados:** Interpretación de los resultados obtenidos.
9. **Revisión de la parte experimental:** Una vez dada por finalizada la parte experimental, se ha hecho una revisión exhaustiva de los códigos y de los resultados obtenidos.
10. **Elaboración de la memoria:** Desarrollo del informe.
11. **Revisión de la memoria:** Una vez terminado el trabajo, se ha hecho una revisión exhaustiva de la memoria.

Téngase en cuenta hay tareas que se han realizado casi simultáneamente, como serían la “Modificación de los algoritmos”, “Obtención de los resultados” y “Análisis de los resultados”. Esto se debe a la necesidad de saber cómo han influido las modificaciones para empezar a estudiar qué otra modificación podría ser beneficiosa. Por ejemplo, si se converge rápidamente a una solución, hay que estudiar por qué ha pasado y, una vez hecha la hipótesis, estudiar qué se podría modificar para que no suceda.

Una estimación del tiempo (en horas) dedicado a cada tarea se puede encontrar en la tabla 2.2.



**Tiempo dedicado**

Actividad	Duración (horas)
Planteamiento y comprensión del problema	20
Búsqueda de información y lecturas	15
Planificación del proyecto	5
Implementación de la propuesta inicial	10
Adaptación de los algoritmos bases	5
Modificación de los algoritmos	200
Obtención de resultados	50
Análisis de los resultados	15
Revisión de la parte experimental	5
Elaboración de la memoria	170
Revisión de la memoria	10
<b>Total</b>	<b>490</b>

Cuadro 2.2: Tiempo dedicado

**2.2. Presupuesto**

Si quisiéramos valorar económicamente el proyecto, tenemos que tener en cuenta dos aspectos fundamentales: el precio de la mano de obra y el de cómputo como si tuviésemos que pagarlo. Sin embargo

El precio de la mano de obra son 25€ la hora. El ordenador portátil usado para la realización de las ejecuciones ha sido un Asus Tuf Gaming A15 FA506IU-HN278 con un procesador AMD® Ryzen™ 7 4800H APU y 16GB (8GB×2) de RAM.

Por lo tanto, el **presupuesto del proyecto queda fijado en 12250€**, a razón de 25€ por cada hora de trabajo dedicada al proyecto.



## Capítulo 3: Repaso bibliográfico



## Capítulo 4: Contexto matemático

Los algoritmos usan operadores matemáticos con el objetivo de alcanzar una solución a los problemas a los que son aplicados y también es necesario obtener el óptimo. La prueba de este hecho puede ser realizada gracias a la teoría de optimización del Análisis Numérico. Nos centraremos en esta teoría, aportando resultados que nos permitan saber si un punto es el óptimo de una función y su relación con los algoritmos básicos de optimización global.

La programación no lineal es un área de las matemáticas aplicadas que involucra problemas de optimización cuando las funciones son no lineales. Nuestro objetivo es introducir este problema y revisar las condiciones generales de optimalidad, que son la base de muchos algoritmos por sus soluciones. Tras esto, aportaremos numerosas nociones relativas al rendimiento de los algoritmos en términos de convergencia, orden de convergencia y comportamiento numérico.

Cabe mencionar que aunque el problema abordado en este trabajo implica la maximización del valor de la solución, se utilizará la notación convencional y usual para tratar este tipo de problemas, es decir, la minimización del valor de la solución. Esto se debe a que, al fin y al cabo, son problemas equivalentes. Una forma sencilla de transformar un problema de maximización en uno de minimización es cambiar el símbolo de los valores. Por ello, en tanto que ambos problemas son equivalentes, se mantendrá la notación de minimizar la función objetivo.

### 4.1. Definición del problema

En primer lugar, daremos una definición a nuestro problema.

**Definición 4.1** Consideramos el problema de calcular el valor de un vector de variables de decisión  $x \in \mathbb{R}^n$  que minimiza la función objetivo  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  donde  $x$  pertenece a un conjunto factible de soluciones  $\mathcal{F} \in \mathbb{R}^n$ . Consideramos el siguiente problema:

$$\min_{x \in \mathcal{F}} f(x) \quad (4.1)$$

**Nota:** Llamaremos **conjunto factible** al espacio de soluciones, es decir, al conjunto de todos los puntos posibles de un problema de optimización que satisface las restricciones del problema.

Ahora, presentamos dos casos de ello:

- El conjunto factible de soluciones  $\mathcal{F}$  es todo el espacio  $\mathbb{R}^n$ . En este caso, el problema es el siguiente:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (4.2)$$

Diremos que el problema 4.1 es no restringido. De forma general, el problema 4.1 es no restringido si  $\mathcal{F}$  es un conjunto abierto.

- El conjunto factible de soluciones está descrito por restricciones de desigualdad y/o igualdad en las variables de decisión:

$$\mathcal{F} = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \dots, p; h_j(x) = 0, j = 1, \dots, m\} \quad (4.3)$$

Entonces, el problema 4.1 se convierte en:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{aligned} \quad (4.4)$$

donde  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  y  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . En este caso, diremos que el problema es restringido.

El problema 4.1 es no lineal cuando al menos una de las funciones del problema es no lineal, es decir,  $f, g_i, i = 1, \dots, p, h_j, j = 1, \dots, m$  es no lineal en  $x$ .

Normalmente, asumimos que en un problema del tipo 4.1 el número de condiciones de igualdad,  $m$ , es menor que el número de variables,  $n$ ; en otro caso, el conjunto factible de soluciones será el vacío, a no ser que haya dependencia en las restricciones. Si solo hay condiciones de igualdad, el problema se llamará “**problema no lineal con restricciones de igualdad**”. Equivalentemente, se tiene el caso de que solo aparezcan condiciones de desigualdad.

En lo siguiente asumiremos que nuestras funciones  $f, g, h$  son diferenciables y continuas en  $\mathbb{R}^n$ . Además, cuando  $f$  sea una función convexa y el conjunto  $\mathcal{F}$  sea también convexo, al problema se le llamará “**problema convexo no lineal**”. Particularmente,  $\mathcal{F}$  es convexo si las funciones que nos aportan las restricciones de desigualdad son convexas y las funciones de las restricciones de igualdad son afines.

La convexidad nos permite añadir estructura a estos problemas y poder explotarlo desde un punto de vista teórico y computacional, esto se debe a que si  $f$  es una función convexa cuadrática y  $g, h$  son afines, entonces tenemos que tratar con un problema de programación cuadrática. Sin embargo, nos centraremos en problemas no lineales generales, sin asumir convexidad.

**Definición 4.2** Un punto  $x^* \in \mathcal{F}$  es un **solución global** de 4.1 si  $f(x^*) \leq f(x), \forall x \in \mathcal{F}$ . El punto es una **solución global estricta** si  $f(x^*) < f(x), \forall x \in \mathcal{F}, x \neq x^*$ .

La existencia de soluciones globales se debe a la compacidad de  $\mathcal{F}$ , en relación con el teorema de Weierstrass:

**Teorema 4.1 (Teorema de Weierstrass)** Sea  $a, b \in \mathbb{R}$  con  $a < b$  y sea  $f : [a, b] \rightarrow \mathbb{R}$  una función continua. Entonces, el intervalo  $f([a, b])$  es cerrado y acotado.

Una consecuencia directa para los problemas sin restricciones es que una solución global existe si el conjunto  $\mathcal{L}^\alpha = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$  es compacto para un  $\alpha$  finito.

**Definición 4.3** Un punto  $x^* \in \mathcal{F}$  es una **solución local** del problema del problema 4.1 si existe  $x^*$  en un vecindario abierto  $\mathcal{B}_{x^*}$  de  $x^*$  tal que  $f(x^*) \leq f(x), \forall x \in \mathcal{F} \cap \mathcal{B}_{x^*}$ . Además, es una **solución local estricta** si  $f(x^*) < f(x), \forall x \in \mathcal{F} \cap \mathcal{B}_{x^*}, x \neq x^*$ .

Determinar una solución global a un problema de este tipo es normalmente una tarea complicada. Los algoritmos que resuelven este tipo de problemas se usan para alcanzar óptimos locales. Incluso en aplicaciones prácticas obtener este tipo de soluciones puede ser también algo bueno.

Ahora introduciremos **notación**:

- Dado un vector  $y \in \mathbb{R}^n$ , definimos su **traspuesta** por  $y'$ . Esta definición puede ser extendida a las matrices  $A \in \mathbb{R}^n \times \mathbb{R}^n$ .
- Dada una función  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , denotamos el vector **gradiente** por  $\nabla h(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$  y la matriz **Hessiana** se denota por

$$\nabla^2 h(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

- Dada una función vector  $w : \mathbb{R}^n \rightarrow \mathbb{R}^q$ , denotamos la matriz de  $n \times q$  cuyas columnas son  $\nabla w_j(x)$ ,  $j = 1, \dots, q$  con  $\nabla w(x)$ .
- Sea  $y$  un vector,  $y \in \mathbb{R}^q$ , denotamos la **norma Euclídea** por  $\|y\|$ . Supongamos que sus componentes son  $y_i$ ,  $i = 1, \dots, q$  y sea  $A$  una matriz cuyas columnas sean  $a_j$ ,  $j = 1, \dots, q$ . Sea  $\mathcal{K} \subset \{1, \dots, q\}$  un subconjunto de índices. Denotamos por  $y_{\mathcal{K}}$  al subvector de  $y$  con componentes  $y_i$  tales que  $i \in \mathcal{K}$  y por  $A_{\mathcal{K}}$  a la submatriz de  $A$  compuesta por las columnas  $a_j$  con  $j \in \mathcal{K}$ .

## 4.2. Condición de Optimalidad

Nuestras soluciones locales deben satisfacer algunas condiciones de optimalidad necesarias. Si nos referimos a problemas como 4.2, tenemos uno de los resultados más conocidos del Análisis Numérico clásico:

**Proposición 4.1** *Sea  $x^*$  una solución local al problema 4.2, entonces*

$$\nabla f(x^*) = 0 \quad (4.5)$$

*Además, si  $f$  es continuamente 2-diferenciable, entonces*

$$y' \nabla^2 f(x^*) y \geq 0, \forall y \in \mathbb{R}^n \quad (4.6)$$

**Nota:** Diremos que una función es continuamente  $i$ -diferenciable si es diferenciable  $i$  veces y dichas diferenciales son continuas.

Si tenemos problemas como 4.1, la mayoría de las condiciones necesarias de optimalidad usadas en el desarrollo de los algoritmos asumen que una solución local debe satisfacer algunas de estas condiciones para evitar alcanzar casos degenerados. Estas condiciones se suelen llamar “**calificaciones de restricción**” y, entre ellas, la más simple y más comúnmente usada es el requisito de restricción de independencia lineal:

**Definición 4.4** *Sea  $\hat{x} \in \mathcal{F}$ . Decimos que la restricción de desigualdad  $g_i$  está **activa** en  $\hat{x}$  si  $g_i(\hat{x}) = 0$ . Denotamos por  $\mathcal{F}_a(\hat{x})$  al conjunto de índices de las restricciones de desigualdad activas en  $\hat{x}$ :*

$$\mathcal{F}_a(\hat{x}) = \{i \in \{1, \dots, p\} : g_i(\hat{x}) = 0\} \quad (4.7)$$

Nótese que en la definición anterior podemos concluir que las restricciones de igualdad  $h_j$  son activas en  $\hat{x}$ . La restricción de independencia lineal se satisface si el gradiente de las restricciones activas son linealmente independientes.

Bajo las condiciones de independencia lineal asumidas, los problemas de restricciones como 4.1 pueden resolverse usando la función Lagrangiana generalizada:

$$L(x, \lambda, \mu) = f(x) + \lambda'g(x) + \mu'h(x) \quad (4.8)$$

donde  $\lambda \in \mathbb{R}^p$ ,  $\mu \in \mathbb{R}^m$  son multiplicadores de Lagrange. El siguiente resultado nos dan información sobre la existencia de estos multiplicadores.

**Proposición 4.2** *Sea  $x^*$  una solución local de 4.1 y supongamos que la independencia lineal se satisface en  $x^*$ . Entonces, los multiplicadores de Lagrange  $\lambda^* \geq 0$ ,  $\mu^*$  existe de tal forma que:*

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0 \quad (4.9)$$

y

$$\lambda'^* g(x^*) = 0$$

Además, si  $f, g, h$  son continuamente 2-diferenciables, entonces:

$$y' \nabla_x^2 L(x^*, \lambda^*, \mu^*) y \geq 0, \forall y \in \mathcal{N}(x^*)$$

donde

$$\mathcal{N}(x^*) = \{y \in \mathbb{R}^n : \nabla g'_{\mathcal{F}_a} y = 0; \nabla h(x^*)' y = 0\}$$

Si  $x \in \mathcal{F}$  satisface una condición de optimalidad suficiente, entonces dicho punto es una solución local al problema 4.1. Para problemas generales, las condiciones de optimalidad suficientes pueden ser establecidas bajo la asunción de que las funciones problemas son continuamente 2-diferenciables, así que tenemos condiciones de suficiencia de segundo orden. Estas condiciones cambian si el problema a tratar es 4.2 o 4.1. Por lo tanto, los siguientes resultados mostrarán cuándo  $x^*$  son óptimos locales en ambos casos.

**Proposición 4.3** *Asumimos que  $x^*$  satisface la condición 4.5. También asumiremos que*

$$y' \nabla^2 f(x^*) y > 0, \forall y \in \mathbb{R}^n, y \neq 0$$

*es decir, se asume que  $\nabla^2 f(x^*)$  es definida positiva. Entonces,  $x^*$  es una solución local estricta del problema 4.2.*

**Proposición 4.4** *Asumimos que  $x^* \in \mathcal{F}$  y  $\lambda^*, \mu^*$  satisfacen las condiciones de 4.9. Asumimos también que*

$$y' \nabla_x^2 L(x^*, \lambda^*, \mu^*) y > 0, \forall y \in \mathcal{P}(x^*), y \neq 0 \quad (4.10)$$

donde

$$\mathcal{P}(x^*) = \{y \in \mathbb{R}^n : \nabla g'_{\mathcal{F}_a} y \leq 0, \nabla h(x^*)' y = 0; \nabla g_i(x^*)' y = 0, i \in \mathcal{F}_a(x^*), \lambda_i^*\}$$

*Entonces,  $x^*$  es una solución local estricta de 4.1.*



Nótese también que  $\mathcal{N}(x^*) \subseteq \mathcal{P}(x^*)$  y la igualdad se da si  $\lambda_i^* > 0, \forall i \in \mathcal{F}_a(x^*)$ .

Una característica importante y principal de los problemas convexos es que una solución global (estricta) del problema es también una solución local (estricta). Además, cuando  $f$  es (estrictamente) convexa, las funciones  $g_i$  son también convexas y las  $h_j$  son afines, entonces las condiciones de optimalidad necesarias dadas en términos de primeras derivadas parciales son suficientes para que un punto  $x^*$  sea una solución global (estricta).

Las condiciones de optimalidad son esenciales para problemas no lineales. Si conocemos la existencia de un óptimo global, entonces el método más común de obtenerlo es el siguiente:

1. Encontrar todos los puntos que satisfacen las condiciones necesarias de primer orden.
2. Tomas el óptimo global como el punto con el valor más bajo dado por la función objetivo
3. Si la función problema es 2-diferenciable, entonces comprueba la condición necesaria de segundo orden y elimina los puntos que no la satisfagan.
4. Para el resto de puntos comprobaremos la condición de suficiencia de segundo orden para encontrar el mínimo local.

Tenemos que destacar que este método no funciona en casos prácticos excepto por casos simples, esto se debe a que tenemos que calcular la solución de un sistema de ecuaciones dado por  $\nabla f(x) = 0$  y este sistema es normalmente no trivial. Entonces, ¿dónde son importantes estas condiciones? Las condiciones de optimalidad son importantes en el desarrollo y análisis de algoritmos.

Un algoritmo que intenta resolver un problema dado por 4.1 genera una secuencia de soluciones factibles  $x^k, k = 0, 1, \dots$  y normalmente finaliza cuando se satisface un criterio de parada. Este criterio se suele basar en la satisfacción de condiciones de optimalidad necesarias dentro de una tolerancia prefijada. Adicionalmente, estas condiciones normalmente sugieren cómo mejorar la solución actual, con lo que la siguiente debería encontrarse más cercana al óptimo.

Así, las condiciones de optimalidad necesarias nos dan la base para el análisis de convergencia de algoritmos. Por lo tanto, las condiciones de suficiencia juegan un papel importante en el análisis del orden de convergencia.

### 4.3. Rendimiento de los algoritmos

#### 4.3.1. Convergencia y Orden de Convergencia

**Definición 4.5** Sea  $\Omega \subset \mathcal{F}$  el subconjunto de puntos que satisfacen las condiciones necesarias de optimalidad del problema 4.1. Un algoritmo se detiene cuando un punto  $x^* \in \Omega$  es calculado. Por lo tanto, a este conjunto se le llamará **conjunto objetivo**.

Un ejemplo de este conjunto para el problema sin restricciones podría ser  $\Omega = \{x \in \mathbb{R}^n : \nabla f(x) = 0\}$ ; mientras que para el problema restringido este conjunto será el conjunto de puntos que satisfacen 4.9.

**Definición 4.6** Sea  $x^k, k = 0, 1, \dots$  la secuencia de puntos generados por un algoritmo. Entonces, el algoritmo es **globalmente convergente** si un punto límite  $x^*$  de  $x^k$  existe tal que  $x^k \in \Omega$  para cualquier punto de inicio  $x^0 \in \mathbb{R}^n$ .

Diremos que el algoritmo es **localmente convergente** si la existencia de  $x^* \in \Omega$  solo puede ser establecida si el punto inicial,  $x^0$ , pertenece a un vecindario de  $\Omega$ .

La definición de convergencia establecida anteriormente es la más débil que asegura que un punto  $x^k$  arbitrariamente cercano a  $\Omega$  puede ser obtenido con un  $k$  suficientemente grande.

En el caso de no tener restricciones, esto implica

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0$$

El requerimiento más fuerte de convergencia nos muestra que la secuencia  $x^k$  converge a un punto  $x^* \in \Omega$ .

Ahora mostraremos cómo se puede definir el orden de convergencia de un algoritmo. Así, podemos asumir por simplicidad que los algoritmos generan una secuencia  $x^k$  que converge a un punto  $x^* \in \Omega$ . El concepto más empleado en términos de convergencia es el Q-orden de convergencia, el cual considera el cociente entre dos iteraciones sucesivas dado por

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|}$$

Entonces, podemos definir el siguiente tipo u orden de convergencia.

**Definición 4.7** El orden de convergencia es Q-lineal si existe una constante  $r \in (0, 1)$  tal que

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq r,$$

para cualquier  $k$  suficientemente grande.

**Definición 4.8** El orden de convergencia es Q-superlineal si

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0$$

**Definición 4.9** El orden de convergencia es Q-cuadrático si

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \leq R$$

para cualquier  $k$  suficientemente grande y donde  $R > 0$  es una constante.

### 4.3.2. Comportamiento numérico

A pesar del rendimiento teórico que puedan tener los algoritmos, otro aspecto importante es el comportamiento práctico. En efecto, si tenemos una gran cantidad de operaciones algebraicas por operación, es posible superar una tasa de convergencia rápida. Tenemos numerosas medidas para evaluar el comportamiento numérico. Sin embargo, si la carga computacional (operaciones algebraicas por iteración) no es despreciable, entonces podemos usar algunas medidas dadas por el número de iteraciones, número de evaluaciones de funciones objetivo, etc.

Medir el rendimiento de los algoritmos es importante para problemas no lineales de gran escala. El término “gran escala” depende de la máquina que se encargue de los datos, pero ese tipo de problema son normalmente problemas sin restricciones que verifican que  $n \geq 1000$ , donde  $n$  es el número de variables. Sin embargo, un problema con restricciones se considerará ser un problema de gran escala cuando el número de variables es  $n \geq 100$  y cuando la suma de las condiciones es 100 o mayor. Uno de los problemas más importantes es el trasladar algoritmos eficientes en problemas a pequeña escala a problemas de gran escala.

## 4.4. Programación no lineal sin restricciones

En esta sección consideraremos algoritmos que traten de resolver el siguiente problema sin restricciones

$$\min_{x \in \mathbb{R}^n} f(x) \quad (4.11)$$

donde  $x \in \mathbb{R}^n$  es el vector de variables de decisión y  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es la función objetivo. Es lógico pensar que si somos capaces de resolver este problema, el procedimiento para ello puede ser ajustado a problemas con restricciones porque solo necesitaremos un conjunto abierto factible de soluciones  $\mathcal{F}$ , y un punto inicial  $x^0 \in \mathcal{F}$ .

Asumimos por simplicidad que nuestra función  $f$  es continuamente 2-diferenciable en  $\mathbb{R}^n$ . Además, asumiremos lo siguiente para asegurarnos de la existencia de una solución de la ecuación 4.11:  $\mathcal{L}^0 = \{x \in \mathbb{R} : f(x) \leq f(x^0)\}$  es compacto para algún  $x^0 \in \mathbb{R}^n$ .

### 4.4.1. Algoritmos de Optimización sin restricciones

Ahora presentaremos algunos modelos de algoritmos que serán usados para resolver los problemas previos. Estos tipos de algoritmos están caracterizados por generar una secuencia de puntos,  $\{x^k\}$ , empezando por un punto inicial  $x^0$ , usando la siguiente iteración

$$x^{k+1} = x^k + \alpha^k d^k \quad (4.12)$$

donde  $d^k$  es la dirección de búsqueda y  $\alpha^k$  es el tamaño de paso junto con  $d^k$ . En este método tenemos dos parámetros a modificar: la dirección de búsqueda y el tamaño del paso, por lo tanto, dependiendo de como variamos estos datos, obtendremos diferentes métodos y esto afectará a las propiedades de convergencia. El tamaño de paso afecta a la convergencia global, mientras que la dirección de búsqueda afecta a la convergencia local.

El siguiente resultado nos da una relación entre convergencia y dirección de búsqueda:

**Proposición 4.5** Sea  $\{x^k\}$  la secuencia generada por 4.12. También asumimos:

1.  $d^k \neq 0$  si  $\nabla f(x^k) \neq 0$ .
2.  $\forall k$  tenemos  $f(x^{k+1}) \leq f(x^k)$ .
- 3.

$$\lim_{k \rightarrow \infty} \frac{\nabla f(x^k)' d^k}{\|d^k\|} = 0 \quad (4.13)$$

4.  $\forall k$  con  $d^k \neq 0$ , tenemos  $\frac{|\nabla f(x^k)' d^k|}{\|d^k\|} \geq c \|\nabla f(x^k)\|$  con  $c > 0$ .

Entonces, tenemos que, o bien, existe  $\hat{k}$  tal que  $x^{\hat{k}} \in \mathcal{L}^0$  y  $\nabla f(x^{\hat{k}}) = 0$ , o bien, se genera una secuencia infinita tal que:

1.  $\{x^k \in \mathcal{L}^0\}$ .
2.  $\{f(x^k)\}$  converge.
- 3.

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0 \quad (4.14)$$

La tercera condición implica que solo necesitamos una subsecuencia que tenga un punto límite en  $\Omega$ . Este resultado nos da información sobre cómo es la convergencia en términos de la dirección de búsqueda.

Procedemos a describir dos métodos conocidos como algoritmos de Búsqueda Lineal y métodos basados en el gradiente.

#### 4.4.2. Algoritmos de Búsqueda Lineal

Estos algoritmos están caracterizados por determinar el tamaño de paso  $\alpha^k$  junto con la dirección de búsqueda  $d^k$ . El objetivo es elegir un tamaño de paso que asegure la convergencia de 4.12. Una elección puede ser elegir dicho tamaño de paso tal y como se describe en la siguiente ecuación:

$$\alpha^k = \arg \min_{\alpha} f(x^k + \alpha d^k)$$

La ecuación anterior puede ser resumida en la siguiente idea: el tamaño de paso es el valor que minimiza la función objetivo junto con una dirección dada. Sin embargo, en esta situación es posible que el mínimo no pueda ser alcanzado porque la función necesita tener propiedades que nos permitan calcular dicho mínimo, por lo tanto, una búsqueda lineal no tiene por qué suponer la mejor solución computacionalmente hablando.

Por ello, tenemos que usar métodos de aproximación. Uno de ellos es calcular el gradiente de la función. En estos casos podemos asumir que

$$\nabla f(x^k)' d^k < 0, \forall k \quad (4.15)$$

Los algoritmos de búsqueda lineal más simples están proporcionados por Armijo.

La elección inicial de  $\nabla^k$  se debe a la dirección  $d^k$  porque garantizamos que, en un número finito de pasos,  $\alpha^k$  tiene un valor tal que  $f(x^{k+1}) < f(x^k)$ , por lo que las condiciones de convergencia de la proposición se encuentran satisfechas.

Esta búsqueda lineal encuentra un tamaño de paso que satisface la condición de decrecimiento suficiente de la función objetivo y, consecuentemente, el desplazamiento suficiente de la secuencia actual.

En algunos casos, necesitamos considerar que los algoritmos de búsqueda lineal no necesitan información sobre las derivadas. En estos casos, la condición 4.15 no puede ser verificada. Por lo tanto, la dirección  $d^k$  no tiene por qué ser descendiente.

Otra posible modificación al algoritmo propuesto podría ser sustituir la condición de parada por la siguiente, que no tiene información sobre la derivada:

$$f(x^k + \alpha d^k) \leq f(x^k) - \gamma \alpha^2 \|d^k\|^2 \quad (4.16)$$

**Algorithm 1** Algoritmo de Búsqueda Lineal de Armijo

---

```

1: procedure (Busqueda_Lineal)( $\delta \in (0, 1), \gamma \in (0, 1/2), c \in (0, 1)$ )
2:   Elegir  $\nabla^k$  tal que
      
$$\nabla^k \geq c \frac{|\nabla f(x^k)' d^k|}{\|d^k\|^2}$$

3:    $\alpha = \nabla^k$ 
4:    $N \leftarrow n$ 
5:   if  $f(x^k + \alpha d^k) \leq f(x^k) + \gamma \alpha \nabla f(x^k)' d^k$  then
6:      $\alpha^k = \alpha$  y parar
7:   else
8:     Establecer  $\alpha = \delta \alpha$  y volver al paso anterior.
9:   end if
10: end procedure

```

---

**4.4.3. Métodos de Gradiente**

El método de gradiente o método del descenso más rápido (*steepest-descent*) se considera un método básico de entre todos los algoritmos de optimización no restringidos. Tiene la regla básica de establecer  $d^k = -\nabla f(x^k)$  en 4.12. Solo necesita información sobre la primera derivada y es importante debido a que el coste computacional y el almacenamiento disponibles son limitados.

Este método es un ejemplo de convergencia global, porque si usamos un algoritmo de búsqueda local adecuado, podemos establecer un resultado de convergencia global. El problema es su orden de convergencia, el cual es bajo y esto es la razón por la que este algoritmo no se suele usar solo. El esquema de este algoritmo es el siguiente:

**Algorithm 2** Método de Gradiente

---

```

1: procedure (Gradiente)
2:   Establecer  $x^0 \in \mathbb{R}^n$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     Establecer  $d^k = -\nabla f(x^k)$  y encuentra un tamaño de paso  $\alpha^k$  usando 1
5:     Establecer  $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$  y  $k = k + 1$ .
6:   end while
7: end procedure

```

---

La elección inicial del tamaño de paso como la dada por Armijo es importante, ya que puede afectar al comportamiento del algoritmo. En términos de convergencia, el siguiente resultado nos asegura la convergencia sin necesitar asumir la compacidad del conjunto  $\mathcal{L}^0$ .

**Nota:** Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .  $f$  se dice **Lipschitziana** si existe  $K > 0$  tal que

$$\|f(x) - f(y)\| \leq K \|x - y\|, \forall x, y \in \mathbb{R}^n$$

**Proposición 4.6** Si  $\nabla f$  es Lipschitziana, continua y  $f$  está acotada inferiormente, entonces la secuencia generada por 2 satisface la tesis de 4.5.

Como conclusión, con este resultados podemos comprobar que el gradiente es bueno en términos de convergencia global. Sin embargo, solo podemos probar la convergencia lineal. Desde un punto de vista práctico, el orden de convergencia es muy pobre y depende del número de condiciones de la matriz Hessiana.

#### 4.4.4. Métodos de Gradiente Conjugado

Este algoritmo es muy popular debido a su simplicidad y sus bajos requerimientos computacionales. En efecto, solo necesita saber sobre las derivadas de primer orden. La idea principal es que la minimización de funciones cuadráticas estrictamente convexas en  $\mathbb{R}^n$  como la siguiente

$$f(x) = \frac{1}{2}x'Qx + \alpha'x \quad (4.17)$$

donde  $Q$  es una matriz simétrica definida positiva, puede ser dividida en  $n$  minimizaciones sobre  $\mathbb{R}$ . Esto se puede hacer utilizando  $n$  direcciones,  $d^0, \dots, d^{n-1}$  conjugadas con respecto de la matriz Hessiana  $Q$ . Junto con cada dirección, se realiza una búsqueda lineal.

El siguiente algoritmo (3) muestra todo lo mencionado anteriormente.

Se debe notar que el valor de  $\alpha^k$  que se utilizará en el tercer paso es el que minimiza la función  $f(x^k + \alpha d^k)$ . Además, el cociente está bien definido porque para dos direcciones distintas tenemos  $(d^j)'Qd^i = 0$ , con  $i, j$  tales que  $i \neq j$ .

---

**Algorithm 3** Algoritmo de direcciones conjugadas para funciones cuadráticas

---

```

1: procedure (Direcciones_Conjugadas) (direcciones Q-conjugadas  $d^0, \dots, d^{n-1}$ )
2:   Establecer  $x^0 \in \mathbb{R}^n$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     Establecer  $\alpha^k = \frac{\nabla f(x^k)'d^k}{(d^k)'Qd^k}$ 
5:     Establecer  $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$  y  $k = k + 1$ .
6:   end while
7: end procedure

```

---

En este algoritmo, las direcciones provienen de los datos, mientras que en el algoritmo del gradiente conjugado se calculan de forma iterativa usando la siguiente regla:

$$d^k = \begin{cases} -\nabla f(x^k) & \text{si } k = 0 \\ -\nabla f(x^k) + \beta^{k-1}d^{k-1} & \text{si } k \geq 1 \end{cases}$$

El escalar  $\beta^k$  se elige para reforzar la conjugación entre las direcciones. La opción más común es la propuesta de Fletcher-Reeves:

$$\beta_{FR} = \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2} \quad (4.18)$$

Sin embargo, la fórmula de Polak-Ribière también puede usarse:

$$\beta_{PR} = \frac{\nabla f(x^{k+1})'(\nabla f(x^{k+1}) - \nabla f(x^k))}{\|\nabla f(x^k)\|^2} \quad (4.19)$$

Ambas fórmulas son iguales en el caso cuadrático y diferentes cuando se trate de cualquier otro caso.

El esquema del gradiente conjugado se presenta como sigue:

**Algorithm 4** Algoritmo del gradiente conjugado

---

```

1: procedure (Gradiente_Conjugado)
2:   Establecer  $x^0 \in \mathbb{R}^n$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     Calcular  $\beta^{k-1}$  usando 4.18 o 4.19 y establecer la dirección usando

```

$$d^k = \begin{cases} -\nabla f(x^k) & \text{si } k = 0 \\ -\nabla f(x^k) + \beta^{k-1}d^{k-1} & \text{si } k \geq 1 \end{cases}$$

```

5:     Encontrar  $\alpha^k$  usando un algoritmo de búsqueda lineal que satisfaga las condiciones de Wolfe.
6:     Establecer  $x^{k+1} = x^k + \alpha^k d^k$  y  $k = k + 1$ .
7:   end while
8: end procedure

```

---

Las condiciones de Wolfe mencionadas en 4 son las siguientes:

$$f(x^k + \alpha d^k) \leq f(x^k) + \gamma \alpha \nabla f(x^k)' d^k \quad (4.20)$$

que es la misma que se usó en la Búsqueda Lineal de Armijo, con la siguiente condición siendo más fuerte

$$|\nabla f(x^k + \alpha d^k)' d^k| \leq \beta |\nabla f(x^k)' d^k| \quad (4.21)$$

donde  $\beta \in (\gamma, 1)$  y  $\gamma$  se encuentra en el mismo intervalo que antes.

La forma más simple de asegurar las propiedades de convergencia global en este método es aplicando reinicios periódicos junto con dirección de descenso más rápido. Aún así, el reinicio puede también suceder si alguno de los términos cuadráticos se pierden y pueden causar, o bien, que el método sea ineficiente, o bien, elecciones de caminos sin sentido.

El mecanismo de reinicio se realiza cada  $n$  iteraciones o si se da la siguiente condición:

$$|\nabla f(x^k)' \nabla f(x^{k+1})| > \delta \|\nabla f(x^{k-1})\|^2$$

con  $0 < \delta < 1$ .

#### 4.4.5. Métodos de Newton

Este método se considera uno de los algoritmos más potentes para resolver problemas de optimización sin restricciones. La aproximación cuadrática de la función objetivo en un vecindario de la solución actual,  $x^k$ , considerada es la siguiente

$$q^k(s) = \frac{1}{2} s' \nabla^2 f(x^k) s + \nabla f(x^k)' s + f(x^k)$$

donde necesitamos saber las derivadas de primer y segundo orden de la función objetivo en la iteración número  $k$ . Este algoritmo también necesita calcular una dirección,  $d_N$ , y en ese caso, se obtiene como un punto estacionario de la aproximación anterior y es la solución del sistema dado por

$$\nabla^2 f(x^k) d_N = -\nabla f(x^k) \quad (4.22)$$

Por lo tanto, la matriz  $\nabla^2 f(x^k)d_N$  es no singular, tiene una inversa y la dirección es  $d_N = -(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$ .

El esquema básico algorítmico está definido por la iteración

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1}\nabla f(x^k) \quad (4.23)$$

La calidad de este método se debe al hecho de que si el punto de partida  $x^0$  está próximo a la solución  $x^*$ , entonces la secuencia de puntos generado por la ecuación anterior converge a  $x^*$  de forma superlineal o cuadrática (si la Hessiana es continuamente Lipschitziana en un vecindario de la solución).

Sin embargo, este método tiene algunas desventajas. Una de ellas es la singularidad de la matriz  $\nabla^2 f$  porque en el caso de ser singular, el método no puede definirse. Otra desventaja está relacionada con el punto inicial,  $x^0$ . Puede ser tal que la secuencia generada por 4.23 no converge, pero puede ocurrir la convergencia a un punto máximo.

Debido a estos hechos, el método de Newton necesita algunos cambios para asegurar la convergencia global a la solución. Un método de Newton convergente debería generar una secuencia de puntos  $\{x^k\}$  con las siguientes características:

- Admite un punto límite.
- Cualquier otro punto límite pertenece a  $\mathcal{L}$  y es un punto estacionario de  $f$ .
- Ningún punto límite es un punto máximo de  $f$ .
- Si  $x^*$  es un punto límite de  $\{x^k\}$  y  $\nabla^2 f(x^*)$  es definida positiva, entonces la convergencia es, al menos, superlineal.

Hay dos enfoques para diseñar un método de Newton convergente globalmente: un enfoque con búsqueda lineal y un enfoque con regiones de confianza.

### Modificaciones de Búsqueda Lineal en el Método de Newton

la adaptación del método a este enfoque es el control del tamaño de paso junto con  $d_N$  tal que 4.23 se convierte en

$$x^{k+1} = x^k - \alpha^k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) \quad (4.24)$$

donde  $\alpha^k$  se elige con una buena búsqueda local. Un ejemplo puede ser inicializar  $\Delta^k = 1$  en 1. Adicionalmente a este cambio, la dirección  $d_N$  puede ser perturbada para asegurar la convergencia global del algoritmo. La forma más fácil de realizar este cambio es usar la dirección del descenso más rápido siempre cuando  $d_N^k$  no satisface alguna de las condiciones de convergencia.

Un posible esquema de dicho algoritmo se presenta a continuación:



**Algorithm 5** Método de Newton con Búsqueda Lineal

---

```

1: procedure (ABLNewton) ( $c_1 > 0, c_2 > 0, p \geq 2, q \geq 3$ )
2:   Establecer  $x^0 \in \mathbb{R}^n$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     if  $\exists d_N^k$  solución de  $\nabla^2 f(x^k) d_N^k = -\nabla f(x^k)$  y satisface
           
$$\nabla f(x^k)' d_N^k \leq -c_1 \|\nabla f(x^k)\|^q, \|d_N^k\|^p \leq c_2 \|\nabla f(x^k)\|$$

       then
5:       Establecer la dirección  $d^k = d_N^k$ 
6:     else
7:        $d^k = -\nabla f(x^k)$ 
8:     end if
9:     Encontrar  $\alpha^k$  usando 1
10:    Establecer  $x^{k+1} = x^k + \alpha^k d^k$  y, después,  $k = k + 1$ 
11:  end while
12: end procedure

```

---

En el tercer paso, se toma la dirección del descenso más rápido si  $\nabla f(x^k)' d_N^k \geq 0$ . Otra posible modificación del método de Newton podría ser aquella que toma la dirección de la curvatura negativa, es decir,  $d^k = -d_N^k$ . Esta modificación se puede hacer si las siguientes dos condiciones se cumplen:

1.  $|\nabla f(x^k)' d^k| \geq c_1 \|\nabla f(x^k)\|^q$

2.  $\|d^k\|^p \leq c_2 \|\nabla f(x^k)\|$

Una segunda modificación es perturbar la matriz Hessiana con una matriz definida positiva  $Y^k$  y ahora la solución provendría de resolver el sistema  $(\nabla^2 f(x^k) + Y^k)d = -\nabla f(x^k)$ .

Las modificaciones comunes de los métodos de Newton se basan en el decremento monótono de los valores de la función objetivo. Con estos cambios la región de convergencia del método puede aumentar más de lo esperado, pero una secuencia convergente en este conjunto puede no ser una secuencia monótonamente descendente de los valores de la función objetivo.

La siguiente modificación está basada en las condiciones de búsqueda lineal y mantiene la propiedad de convergencia global. En esta modificación, usaremos una regla no monótona. Por lo tanto, el método obtenido es el siguiente:

**Algorithm 6** Método de Newton no monótono

---

```

1: procedure (NewtonNM) ( $c_1 > 0, c_2 > 0, p \geq 2, q \geq 3$  y  $M$  entero)
2:   Establecer  $x^0 \in \mathbb{R}^n$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     if  $\exists d_N^k$  solución de  $\nabla^2 f(x^k) d_N^k = -\nabla f(x^k)$  y satisface
           
$$\nabla f(x^k)' d_N^k \leq -c_1 \|\nabla f(x^k)\|^q, \|d_N^k\|^p \leq c_2 \|\nabla f(x^k)\|$$

           then
5:       Establecer la dirección  $d^k = d_N^k$ 
6:     else
7:        $d^k = -\nabla f(x^k)$ 
8:     end if
9:     Encontrar  $\alpha^k$  usando 1, tal que
           
$$f(x^k + \alpha d^k) \leq \max_{0 \leq j \leq J} \{f(x^{k-j})\} + \gamma \alpha \nabla f(x^k)' d^k$$

           con  $J = m(k)$ 
10:    Establecer  $x^{k+1} = x^k + \alpha^k d^k$  y, después,  $k = k + 1$ 
11:    Establecer  $m(k) = \min\{m(k-1) + 1, M\}$ 
12:  end while
13: end procedure

```

---

Los métodos de Búsqueda Lineal estudiados hasta ahora convergen a puntos satisfaciendo solo las condiciones de optimalidad de primer orden necesarias de la ecuación 4.11. Esto se debe a que el método de Newton no explota toda la información obtenida en la segunda derivada. Es posible obtener una convergencia más fuerte si usamos el par de direcciones  $(d^k, s^k)$  y una búsqueda curvilinear, es decir,

$$x^{k+1} = x^k + \alpha^k d^k + (\alpha^k)^{\frac{1}{2}} s^k \quad (4.25)$$

donde  $d^k$  es una dirección del método de Newton y  $s^k$  es una dirección que incluye información de curvatura negativa con respecto a  $\nabla^2 f(x^k)$ . Con esta idea y algunas modificaciones a la Búsqueda Lineal de Armijo, se pueden crear algoritmos globalmente convergente que además puedan satisfacer las condiciones de optimalidad de segundo orden necesarias para la ecuación 4.11.

**Modificaciones de Regiones de Confianza en el Método de Newton**

Este tipo de algoritmos tienen su iteración principal como muestra la siguiente ecuación

$$x^{k+1} = x^k + s^k$$

donde el paso  $s^k$  se obtiene minimizando la forma cuadrática  $q^k$  de la función objetivo en una región de confianza del espacio  $\mathbb{R}^n$ . La región de confianza se define como una norma  $l_p$  del paso  $s$ . Lo más común es elegir la norma Euclidiana, con la cual, en cada iteración,  $s^k$  es la solución de

$$\min_{s \in \mathbb{R}^n} \frac{1}{2} s' \nabla^2 f(x^k) s + \nabla f(x^k)' s$$

donde  $\|s\|^2 \leq (a^k)^2$  con  $a$  siendo el radio de la región de confianza. Otra opción consistiría en realizar un cambio de escala a la condición previa de la siguiente forma:  $\|D^k s\|^2 \leq (a^k)^2$ . Por simplicidad, asumiremos que la matriz  $D^k = I$ , es decir, la matriz identidad en el espacio adecuado.

Estos algoritmos se caracterizan por la siguiente idea: cuando la matriz  $\nabla^2 f(x^k)$  es definida positiva, entonces el radio  $a^k$  tiene que ser lo suficientemente grande que el minimizador de 4.4.5 no tenga restricciones y el paso dado por Newton sea un entero. Además,  $a^k$  se actualiza en cada iteración y su instrucción de actualización depende de la proporción  $\rho^k$  entre la reducción de los valores de la función objetivo  $f(x^k) - f(x^{k+1})$  y la reducción esperada  $f(x^k) - q^k(s^k)$ .

El siguiente algoritmo presenta las ideas anteriores y también garantiza la satisfacción de las condiciones de optimalidad necesarias para 4.11 en su tercer paso.

---

**Algorithm 7** Método de Newton basado en Regiones de Confianza
 

---

```

1: procedure (NewtonRegionConfianza) ( $0 < \gamma_1 \leq \gamma_2 < 1, 0\delta_1 < 1 \leq \delta_2$ )
2:   Establecer  $x^0 \in \mathbb{R}^n$ ,  $k = 0$  y  $a^0 = 0$ .
3:   Encontrar  $s^k = \arg \min_{\|s\| \leq a^k} q^k(s)$ 
4:   if  $f(x^k) == q^k(s^k)$  then
5:     Parar
6:   else
7:     Calcular la proporción

```

$$\rho^k = \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - q^k(s^k)}$$

```

8:     if  $\rho^k \geq \gamma_1$  then
9:       Actualizar  $x^{k+1} = x^k + s^k$ 
10:    else
11:      Actualizar  $x^{k+1} = x^k$ 
12:    end if
13:    Actualizar el radio  $a^k$ 

```

$$a^{k+1} = \begin{cases} \delta_1 a^k & \text{si } \rho^k < \gamma_1 \\ a^k & \text{si } \rho^k \in [\gamma_1, \gamma_2] \\ \delta_2 a^k & \text{si } \rho^k > \gamma_2 \end{cases}$$

y establecer  $k = k + 1$ , entonces volver al paso 3.

```

14: end if
15: end procedure

```

---

Si  $f$  es además continuamente 2-diferenciable, entonces la secuencia del algoritmo anterior,  $\{x^k\}$ , tiene un punto límite que satisface las condiciones de optimalidad necesarias de primer y segundo orden para 4.11. Además, si  $\{x^k\}$  converge a un punto en el que la matriz Hessiana  $\nabla^2 f$  es definida positiva, entonces el orden de convergencia es superlineal.

El esfuerzo computacional de este algoritmo es el subproblema de las regiones de confianza. Debido a este hecho, se han ido desarrollando cada vez más algoritmos para resolverlo. Sin embargo, no necesitamos una solución exacta para esta ecuación. Para probar que la convergencia global del algoritmo es suficiente verificar que el valor  $q^k(s^k)$  es menor que el valor en un punto de Cauchy, que es el punto que minimiza el modelo cuadrático en la región de confianza.

### Métodos de Newton Truncados

Los métodos de Newton necesitan calcular la solución de un sistema lineal de ecuaciones en cada iteración. Si echamos un vistazo a problemas a gran escala, resolver este sistema en cada iteración puede resultar demasiada carga computacionalmente hablando. Además, la solución exacta cuando  $x^k$  está lejos de una solución y  $\|\nabla f(x^k)\|$  es grande no es necesaria. Debido a este hecho, se han propuesto numerosos métodos que calculan una solución aproximada a este sistema con un orden de convergencia bueno, es decir, si  $\widetilde{d}_N^k$  es una solución aproximada del sistema, entonces la medida de precisión es dada por el residual de la ecuación de Newton, que sería

$$r^k = \nabla^2 f(x^k) \widetilde{d}_N^k + \nabla f(x^k)$$

Si podemos controlar este residual, entonces podemos probar la convergencia superlineal.

**Proposición 4.7** *Si  $\{x^k\}$  converge a una solución y si*

$$\lim_{k \rightarrow \infty} \frac{\|r^k\|}{\|\nabla f(x^k)\|} = 0$$

*entonces  $\{x^k\}$  converge de forma superlineal.*

Estos métodos solo requieren de operaciones matriciales-vectoriales, por lo que son adecuados para problemas a gran escala. Al siguiente algoritmo se le conoce como el método de Newton truncado y se necesita que la matriz Hessiana sea definida positiva. Este método se obtiene aplicando un esquema de gradiente conjugado para encontrar una solución óptima del sistema 4.22.

Generaremos los vectores  $d_N^i$ , que se van a ir aproximando a la dirección  $d_N^k$  y se detendrá cuando ocurra uno de los siguientes casos:

- El residual  $r^i$  verifica que  $\|r^i\| \leq \epsilon^k$ , para  $\epsilon^k > 0$ .
- Si se ha encontrado una dirección de curvatura negativa, es decir, la dirección conjugada  $s^i$  verifica que  $(s^i)' \nabla^2 f(x^k) s^i \leq 0$ .

Este algoritmo tiene el mismo resultado para convergencia que el método de Newton. Por simplicidad, eliminaremos las dependencias en la iteración  $k$  y estableceremos  $H = \nabla^2 f(x^k)$  y  $g = \nabla f(x^k)$ . Este método genera las direcciones conjugadas  $s^i$  y los vectores  $p^i$  que aproximan la solución del sistema de Newton  $\widetilde{d}_N^k$

**Algorithm 8** Algoritmo de Newton Truncado

- 
- 1: **procedure** (NewtonRegionConfianza) ( $k, H, g$  y  $\eta > 0$  escalar)  
 2:     Establecer  $i = 0, p^0 = 0, r^0 = -g, s^0 = r^0$  y

$$\epsilon = \eta \|g\| \min \left\{ \frac{1}{k+1}, \|g\| \right\}$$

- 3:     **while do**  
 4:     **end while**  
 5:

$$a_N = \begin{cases} \delta_1 a^k & si \quad \rho^k < \gamma_1 \\ a^k & si \quad \rho^k \in [\gamma_1, \gamma_2] \\ \delta_2 a^k & si \quad \rho^k > \gamma_2 \end{cases}$$

- 6: **end procedure**
- 

Si aplicamos este algoritmo para encontrar una solución aproximada al sistema en el cuarto paso de 5 o 6, entonces obtendremos la versión monótona truncada del método de Newton.

**Métodos Quasi-Newton**

Estos métodos han sido introducidos con el objetivo de diseñar algoritmos eficientes que no necesiten la evaluación de derivadas de segundo orden. Por tanto, han establecido  $d^k$  como la solución de

$$B^k d = -\nabla f(x^k) \quad (4.26)$$

donde  $B^k$  es una matriz definida positiva simétrica de tamaño  $n \times n$  que se ajusta de forma iterativa para que la dirección  $d^k$  tienda a aproximar la dirección del método de Newton. A la fórmula anterior se le conoce como **fórmula quasi-Newton** y su inversa es

$$d^k = -H^k \nabla f(x^k) \quad (4.27)$$

Ambas matrices se modifican en cada iteración como una corrección de la anterior, es decir,  $B^{k+1} = B^k + \Delta B^k$ , y lo mismo pasa para  $H^k$ . Definimos las siguientes dos cantidades:

$$\delta^k = x^{k+1} - x^k \quad \gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

En caso de que  $f$  sea una función cuadrática, entonces la ecuación quasi-Newton sería

$$\nabla^2 f(x^k) \delta^k = \gamma^k \quad (4.28)$$

por lo tanto,  $\Delta B^k$  (lo mismo para  $\Delta H^k$ ) se elige de la siguiente forma

$$(B^k + \Delta B^k) \delta^k = \gamma^k \quad (4.29)$$

La regla de actualización de  $H^k$  está dada por

$$\Delta H = \frac{\delta^k (\delta^k)' }{(\delta^k)' \gamma^k} - \frac{H^k \gamma^k (H^k \gamma^k)'}{(\gamma^k)' H^k \gamma^k} + c(\gamma^k)' H^k \gamma^k v^k (v^k)' \quad (4.30)$$

donde

$$v^k = \frac{\delta^k}{(\delta^k)' \gamma^k} - \frac{H^k \gamma^k}{(\gamma^k)' H^k \gamma^k}$$

y  $c > 0$  es un escalar. El algoritmo que vamos a mostrar ahora fue creado por Broyden, Fletcher, Goldfarb y Shanno. Es un método de búsqueda lineal en el que el tamaño de paso es  $\alpha^k$  se obtiene mediante un algoritmo de búsqueda lineal. El esquema del algoritmo se presenta debajo:

---

**Algorithm 9** Algoritmo BFGS Inverso Quasi-Newton

---

```

1: procedure (InversaBFGSQuasiNewton)
2:   Establecer  $x_0 \in \mathbb{R}^n$ ,  $H^0 = I$  y  $k = 0$ 
3:   while  $\nabla f(x^k) \neq 0$  do
4:     Establecer la dirección  $d^k = -H^k \nabla f(x^k)$ 
5:     Encontrar  $\alpha^k$  mediante una búsqueda lineal que satisfaga las condiciones de Wolfe 4.20
       y 4.21
6:     Actualizar

$$\begin{aligned} x^{k+1} &= x^k + \alpha^k d^k \\ H^{k+1} &= H^k + \Delta H^k \end{aligned} \tag{4.31}$$

       con  $\Delta H^k$  dada por 4.30 con  $c = 1$ .
7:     Establecer  $k = k + 1$ 
8:   end while
9: end procedure

```

---

Distinguimos dos casos en relación a las propiedades de convergencia: caso convexo y caso no convexo. Cuando no tenemos convexidad, si existe una constante  $\rho$  tal que para cada  $k$  se verifica la siguiente condición

$$\frac{\|\gamma^k\|^2}{(\gamma^k)' \delta^k} \leq \rho \tag{4.32}$$

entonces la secuencia de puntos generada por el algoritmo satisface

$$\lim_{k \rightarrow \infty} \inf \|\nabla f(x^k)\| = 0$$

que es la condición débil de ???. Para el caso convexo, la desigualdad anterior se mantiene. El siguiente resultado nos dará información sobre el orden de convergencia de este algoritmo.

**Proposición 4.8** Sea  $\{B^k\}$  una secuencia de matrices no singulares y sea  $\{x^k\}$  la secuencia dada por

$$x^{k+1} = x^k - (B^k)^{-1} \nabla f(x^k)$$

También supondremos que  $\{x^k\}$  converge a un punto  $x^*$  donde  $\nabla^2 f(x^*)$  es también no singular. Entonces, la secuencia  $\{x^k\}$  converge de forma superlineal a  $x^* \Leftrightarrow$

$$\lim_{k \rightarrow \infty} \frac{\|[B^k - \nabla^2 f(x^*)](x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} = 0$$

Este algoritmo está pensado para problemas de pequeña escala, ya que para los de gran escala, el almacenar una matriz como serían  $B^k$  o  $H^k$  ocasionaría problemas de almacenamiento. Por lo tanto, para esos problemas la información se obtiene de las últimas iteraciones.

## Métodos sin derivadas

Estos métodos no calculan explícitamente las derivadas de  $f$ . Son adecuados para cuando, o bien, el gradiente de la función objetivo no puede ser calculado, o bien, cuando es computacionalmente costoso. Sin embargo, si queremos probar las propiedades de convergencia, necesitaremos suponer que  $f$  es continuamente diferenciable.

De entre todos estos algoritmos, los dos más importantes son los algoritmos de búsqueda de patrones (PSA, *pattern-search algorithm*) y los algoritmos de búsqueda lineal sin derivadas (DFLSA, *derivative-free line-search algorithm*). Estos algoritmos son similares, y su diferencia reside en las suposiciones que hacen sobre el conjunto de direcciones y en la regla que usan para encontrar el tamaño de paso junto con las direcciones. Denotamos  $\mathcal{D}^k = \{d^1, \dots, d^r\}$  con  $r \geq n + 1$  como el conjunto de direcciones y suponemos que son unitarias.

También asumimos el siguiente hecho para los PSA: las direcciones  $d^j \in \mathcal{D}^k$  son la  $j$ -ésima columna de la matriz  $B\Gamma^k$  con  $B$  una matriz no singular con coeficientes reales de tamaño  $n$  y  $\Gamma^k \in \mathcal{M} \subset \mathbb{Z}^{n \times r}$ , donde  $\mathcal{M}$  es un conjunto finito de matrices integrales tales que su rango coincide con el número de filas que tienen (*full row-rank*).

Esta suposición nos da una idea para entender que el PSA itera sobre  $x^{k+1}$  en una red (*lattice*) racional centrada en  $x^k$ . Sea  $\mathcal{P}^k$  el conjunto de candidatos para la siguiente iteración, es decir,

$$\mathcal{P} = \{x^{k+1} : x^k + \alpha^k d^j, d^j \in \mathcal{D}^k\}$$

En este conjunto se conoce el patrón y se elige el tamaño de paso para preservar la estructura algebraica en la siguiente iteración, por lo que tenemos  $f(x^k + s^k) < f(x^k)$  con  $s^k = \alpha^k d^j$ .

Para DFLSA haremos la siguiente suposición: las direcciones  $d^j \in \mathcal{D}^k$  son la  $j$ -ésima columna de una matriz  $B^k$  de tamaño  $n \times r$  con rango  $n$ . En este caso, no hay suposiciones adicionales y solo tiene que verificar la reducción de la función objetivo.

Los dos siguientes algoritmos muestran una versión del PSA y del DFLSA.

---

### Algorithm 10 Algoritmo de Búsqueda de Patrones

---

- 1: **procedure** (PSA) ( $\mathcal{D}^{\parallel}$  satisfaciendo la suposición previa,  $\tau \in \{1, 2\}, \theta = \frac{1}{2}$ )
  - 2:   Establecer  $x_0 \in \mathbb{R}^n, \Delta^0 > 0$  y  $k = 0$
  - 3:   Comprobar la convergencia
  - 4:   **if**  $\exists j \in \{1, \dots, r\}$  :
 
$$f(x^k + \alpha^k d^j) < f(x^k), \alpha^k = \Delta^k$$
  - then**
  - 5:     Establecer  $x^{k+1} = x^k, \Delta^{k+1} = \tau \Delta^k, k = k + 1$  e ir al paso 3.
  - 6:   **else**
  - 7:     Establecer  $x^{k+1} = x^k, \Delta^{k+1} = \theta \Delta^k, k = k + 1$
  - 8:   **end if**
  - 9: **end procedure**
-

**Algorithm 11** Algoritmo de Búsqueda Lineal sin Derivadas

---

```

1: procedure (DFLSA) ( $\mathcal{D}^{\parallel}$  satisfaciendo la suposición previa,  $\gamma > 0$ ,  $\theta \in (0, 1)$ )
2:   Establecer  $x_0 \in \mathbb{R}^n$ ,  $\Delta^0 > 0$  y  $k = 0$ 
3:   Comprobar la convergencia
4:   if  $\exists j \in \{1, \dots, r\}$  :
       
$$f(x^k + \alpha^k d^j) \leq f(x^k) - \gamma(\alpha^k)^2, \alpha^k \geq \Delta^k$$

     then
5:     Establecer  $x^{k+1} = x^k$ ,  $\Delta_{k+1} = \alpha^k$ ,  $k = k + 1$  e ir al paso 3.
6:   else
7:     Establecer  $x^{k+1} = x^k$ ,  $\Delta_{k+1} = \theta \Delta^k$ ,  $k = k + 1$ 
8:   end if
9: end procedure

```

---

Ambos algoritmos generan una secuencia que verifica la condición débil de convergencia de 4.5, es decir,

$$\lim_{k \rightarrow \infty} \inf \|\nabla f(x^k)\| = 0$$

En este caso, tomamos un índice,  $j$ , en PSA tal que

$$f(x^k + \alpha^k d^j) = \min_{i: d^i \in \mathcal{D}^k} f(x^k + \alpha^k d^i) < f(x^k)$$

también mantenga la premisa de 4.5. Esto también ocurre en DFLSA, porque solo tenemos que tomar un tamaño de paso,  $\alpha^k$ , tal que

$$f\left(x^k + \frac{\alpha^k}{\delta} d^k\right) \geq \max \left\{ f(x^k + \alpha^k d^k), f(x^k) - \gamma \left(\frac{\alpha^k}{\delta}\right)^2 \right\}, \delta \in (0, 1)$$

Ambos esquemas encuentran un tamaño de paso que les permita comprobar la convergencia del algoritmo. En estos algoritmos no necesitamos tener el gradiente de  $f$ , por lo que tenemos que comprobar la siguiente condición:

$$\sqrt{\frac{\sum_{i=1}^{n+1} (f(x^i) - \bar{f})^2}{n+1}} \leq \text{tolerancia} \quad (4.33)$$

donde  $\bar{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f(x^i)$  y  $\{x^i : i = 1, \dots, n+1\}$  incluye el punto actual y los  $n$  puntos anteriormente generados junto con las  $n$  direcciones.

Como resultado de esta teoría, tenemos algoritmos que nos permitan generar secuencias e puntos en  $\mathbb{R}^n$  que converjan a los puntos óptimos y, bajo ciertas circunstancias, pueden ser el óptimo global. Estos métodos son el principio de la gran cantidad de algoritmos presentados en la literatura. Estos algoritmos proponen formas de alcanzar un óptimo global de funciones basadas en una idea inspirada en la naturaleza.

Incluso si la gran cantidad de algoritmos que presentaremos en las próximas secciones, el diseño de cada uno es similar a cómo esta teoría propone el movimiento, es decir, metaheurísticas basando su movimiento en una dirección dada por un vector y el criterio de parada basado en condiciones que normalmente presentan la optimalidad del mejor individuo de la población .



# Problema y Diseño Experimental



## Algoritmos de Referencia



## Componentes de la propuesta



## Parte Experimental

- 4.5. Algoritmos de Referencia experimentación
- 4.6. Resultados versión expensive
- 4.7. Incorporación del histórico
- 4.8. Uso de GRASP
- 4.9. Operador de Cruce Intensivo
- 4.10. Estudio de la diversidad
- 4.11. Incrementando la diversidad (con nuevo reemplazo)





## Conclusiones



