

Tarea 5 - Análisis de Resultados

1. Ejercicio 1. Parametrización de Algoritmos

(a) Comparar dos esquemas de enfriamiento:

El que ya teníamos era el enfriamiento geométrico y agregamos el enfriamiento lineal, esto se hizo agregando un parámetro para el enfriamiento y al momento de enfriar llamar una función que nos dará el nuevo valor de la temperatura dependiendo de la iteración y el enfriamiento seleccionado.

La comparación la realizaremos en dimensión 10, que fue la que se nos pidió en clase.

Enfriamiento lineal:

$$T_{k+1} = T_0 - \eta k$$

Adicionalmente para el enfriamiento lineal la temperatura se mantiene en múltiples iteraciones, entonces necesitamos saber cuántas iteraciones vamos a mantener la temperatura.

Enfriamiento geométrico:

$$T_{k+1} = \alpha T_k, \alpha \in (0, 1)$$

En nuestro caso para el enfriamiento geométrico usamos $\alpha = 0.999$, para el enfriamiento lineal usamos $\eta = 0.005$ y el número de iteraciones que duraría el enfriamiento lineal sin modificar la temperatura fue de 50.

```
private double enfriar(double temp, int enfriamiento, int numIt) {  
    // Enfriamiento geométrico  
    if (enfriamiento == 1) {  
        return temp * 0.999;  
    }  
    // Enfriamiento lineal  
    if (numIt % 50 == 0) {  
        return temp - 0.005 * (numIt);  
    }  
    return temp;  
}
```

Figure 1: Enfriamientos

Table 1: Comparación de resultados para diferentes enfriamientos

Función	Enfriamiento geométrico		Enfriamiento lineal	
	Mejor	Promedio	Mejor	Promedio
Sphere	2.19×10^{-12}	2.19×10^{-12}	7.651796182	9.993936606
Ackley	7.50×10^{-11}	7.50×10^{-11}	20.14154387	21.12303373
Griewank	0.02961692	0.328792463	21.12298534	21.64817231
Rastrigin	0.469359761	0.956771539	8.355959187	9.857346543
Rosenbrock	1.81×10^{-5}	0.182658631	2.951638085	4.505204739

Los parámetros fueron elegidos a prueba y error, iniciando con valores sugeridos en clase.

Un punto importante a destacar es que para lograr los resultados presentados en la tabla usando el esquema de enfriamiento lineal, fue necesario probar con una basta cantidad de combinaciones de parámetros y en general la mayoría daban peores resultados.

Por otro lado con el esquema de enfriamiento geométrico (que ya habíamos implementado desde la tarea 3) nos facilitó encontrar un parámetro que diera esos, ya que solo fue estar jugando con un solo parámetro.

Sobre los resultados claramente el enfriamiento geométrico superó por mucho el enfriamiento lineal. De hecho podemos notar que en la función Sphere y en la función Ackley el valor promedio y el mejor valor son exactamente iguales. Esto nos podría indicar que dichas funciones convergieron al óptimo global (representable por la partición) con el esquema de enfriamiento geométrico.

En conclusión para nosotros el esquema de enfriamiento geométrico fue mejor, ya que buscar a prueba y error un parámetro que de mejores resultados (respecto al mismo esquema pero con diferente parámetro) fue relativamente fácil, además de que comparado con el enfriamiento lineal dio mejor resultados.

(b) **Comparar los siguientes esquemas de reemplazo:**

i) Generacional

El reemplazo generacional consiste en reemplazar a toda la población actual por la población generada tras los distintos operadores de cruz y recombinación.

La mayor desventaja del reemplazo generacional es que se pueden perder mejores individuos previamente encontrados al momento de avanzar en una generación, ya que la única forma de que un individuo prevalezca usando este esquema de reemplazo (y nuestra selección de operadores) es que sea elegido dos veces en la misma selección de padres de la ruleta y los hijos generados no sean mutados.

ii) Generacional con elitismo

El elitismo significa seleccionar el individuo más apto y pasarlo a la siguiente generación sin modificación alguna, con la finalidad de preservar la mejor solución hasta el momento. [?]

De acuerdo a la investigación hecha, se menciona que el elitismo favorece una convergencia más rápida, preservación de la calidad y eficiencia.[?]

En particular estas tres características van de la mano, ya que una convergencia más rápida nos llevaría a necesitar menos operaciones para llegar al valor al que converge, lo que mejora la eficiencia del algoritmo por la cantidad de generaciones creadas.

iii) Reemplazo de los peores Este esquema de reemplazo es que el tiene mayor diferencia a los dos anteriores.

Lo que nosotros hicimos fue calcular la aptitud promedio de la población y para generar la siguiente generación iríamos reemplazando únicamente a los individuos que tengan una aptitud peor (para minimización es una evaluación mayor) a la promedio.

Este método nosotros lo consideramos que puede ser equivalente al elitista, porque en cada generación nos aseguramos de mantener al menos al mejor individuo, la diferencia con el elitista es que este esquema penaliza aquellos genes cuya aptitud sea la peor, la forma en la que lo hace es reemplazándolos por nuevos hijos.

Si tomáramos como los padres a estos individuos con una aptitud mejor a la promedio, se esperaría que surgiera una convergencia prematura, debido a la reducción de posibles padres, lo que genera menor variabilidad de genes.

Es por eso que al seleccionar los padres para reemplazar a los peores, usamos la ruleta sobre la población entera (con todo y los peores), de modo que los peores individuos aún pueden ser padres. Aunque su cuantil sea menor y tengan menos probabilidad de ser seleccionado como padre; sabemos que la ruleta mantiene el promedio, por lo que la generación de los nuevos hijos es igual al los dos esquemas anteriores. Lo que cambia es que reemplazaremos a los peores.

El elitista pasa directamente a los mejores individuos y el reemplazo de peores de forma indirecta también pasa sin modificaciones a los mejores individuos, además los dos generar de igual forma a los hijos (ruleta sobre la población entera), entonces estos dos esquemas pueden llegar a ser equivalentes.

Una posible ventaja que se nos ocurre del reemplazo de los peores así como lo implementamos es que el "elitismo equivalente" no es fijo, va variando según varía el promedio de evaluación de la población, es decir, la cantidad de individuos a reemplazar y la cantidad de individuos a mantener no va a ser constante.

Table 2: Comparación de resultados para diferentes estrategias de selección y algoritmo genético

Función	Generacional		Elitista		Reemplazo de peores	
	Mejor	Promedio	Mejor	Promedio	Mejor	Promedio
Sphere	0.433	1.258	0.094	0.787	0.542	1.576
Ackley	0.181	2.008	0.063	0.888	0.323	3.241
Griewank	15.304	20.154	13.405	19.103	18.716	21.052
Rastrigin	1.238	2.358	1.687	2.421	1.299	2.344
Rosenbrock	0.515	2.305	0.606	2.591	0.580	2.548

(c) Mejor parametrización

a) Recocido simulado vs Algoritmo genético

Table 3: Comparación de GA con reemplazo elitista y SA con enfriamiento geométrico

Función	Elitista		Enfriamiento Geométrico	
	Mejor	Promedio	Mejor	Promedio
Sphere	0.094	0.787	2.19×10^{-12}	2.19×10^{-12}
Ackley	0.063	0.888	7.50×10^{-11}	7.50×10^{-11}
Griewank	13.405	19.103	0.02961692	0.328792463
Rastrigin	1.687	2.421	0.469359761	0.956771539
Rosenbrock	0.606	2.591	1.81×10^{-5}	0.182658631

Como podemos notar, en todas las filas el recocido simulado obtiene una mejor minimización que el algoritmo genético.

Las ventajas del recocido simulado es que es un algoritmo muy rápido y no es tan pesado computacionalmente como el algoritmo genético que va guardando poblaciones en cada generación. El recocido simulado va trabajando sobre una sola solución activa.

En realidad para este inciso, comparar estas dos metaheurísticas hacerlo solo en dimensión 10 no se me hace una comparación justa, por lo que más adelante usaremos dimensiones más grandes y veremos que sucede.

Para este inciso podríamos decir que parece que el algoritmo genético es peor, ya que en dimensión 10 dio peores resultados que el recocido simulado, tardó más en ejecutarse, es más complicado de implementar. Mientras que el recocido simulado es fácil de implementar.

2. Experimentación en Optimización continua

(a) Para los incisos anteriores ya se habían realizado 30 ejecuciones para cada función y para cada esquema. Se hizo en dimensión 10, pero notamos que usar dimensión 10 no nos estaba ayudando a hacer una comparación justa ya que parecía que el algoritmo genético no tenía sentido, puesto que fue más difícil de implementar, más tardado y dio peores resultados.

Por lo que para la experimentación decidimos usar dimensión 20, tanto para GA elitista como para SA geométrico realizamos 30 ejecuciones por función, con dimensión 20.

(b) **Fijar Criterio de término**

En realidad nuestro algoritmo genético lo teníamos con 1000 generaciones y el recocido simulado en 15000 iteraciones.

De esta forma ambos algoritmo para la misma dimensión se ejecutaban en tiempos muy similares.

Pero para poder generar gráficas más entendibles, corrimos ambos con un criterio de término de 15k iteraciones (teniendo en cuenta que el algoritmo genético desde antes de las 1000 ya había convergido en todas las funciones).

Estas 14k generaciones extra hacían que el algoritmo genético tardara hasta 23 minutos realizando todas las ejecuciones para el algoritmo genético. Ya que 14k generaciones para cada una de las 30 ejecuciones por función (5 funciones), nos hizo computar $1500 * 30 * 5 = 2,250,000$ generaciones extra que no aportaban nada, en comparación a las $1000 * 30 * 5 = 150,000$ que realizaba con las mil generaciones. Aunque tardara mucho más, estas 14k generaciones extra ya no ayudaban a encontrar un mejor resultado.

Por lo que decidimos que era equivalente comparar ambas con 15k iteraciones.

- (c) El programa imprime como salida la semilla del generador de números aleatorios utilizada para la ejecución correctamente.

```
Solución: [-10.133030446298221, 0.07596
78141, 0.26822096543810403, -8.09083177
, 0.8982183690591228, -0.34174927276353

De las 30 ejecuciones:
Mejor valor: 0.7991416860171572
Promedio: 1.2004222983293429
Semillas: [1456-1486]
Dimensión: 20
```

- (d) El programa permite ingresar la semilla del generador de aleatorios para la ejecución correctamente

```
200
Ingrese la semilla para generar la población inicial:
1456
Ingrese la probabilidad de cruce:
```

3. Análisis de resultados

- (a) Tabla de configuración de parámetros para los diferentes algoritmos (o estrategias) con las que se haya trabajado.

Para hacer las comparaciones en el algoritmo genético del primer ejercicio, lo único que variamos entre los parámetros de los 3 fue el esquema de reemplazo, para que fuera una comparación justa y los demás operadores no afectaran nuestros resultados.

Table 4: Configuración del algoritmo genético (GA)

Estrategia	Configuración					
	Op reemplazo	Op cruza	Op mutación	Tam pobl	Rango Seed	Num generaciones
GA	Generacional	n puntos	flip	200	[1456-1485]	1000
GA	Elitista	n puntos	flip	200	[1456-1485]	1000
GA	Peores	n puntos	flip	200	[1456-1485]	1000

Para el recocido simulado fue lo siguiente:

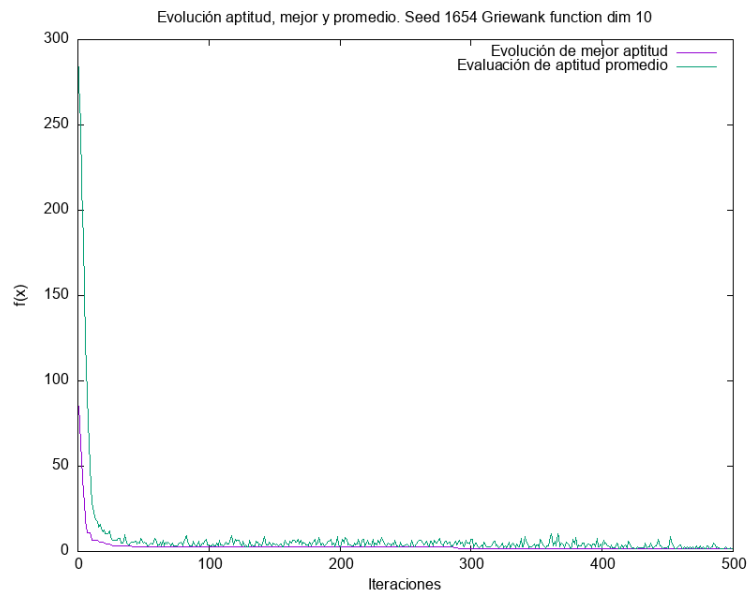
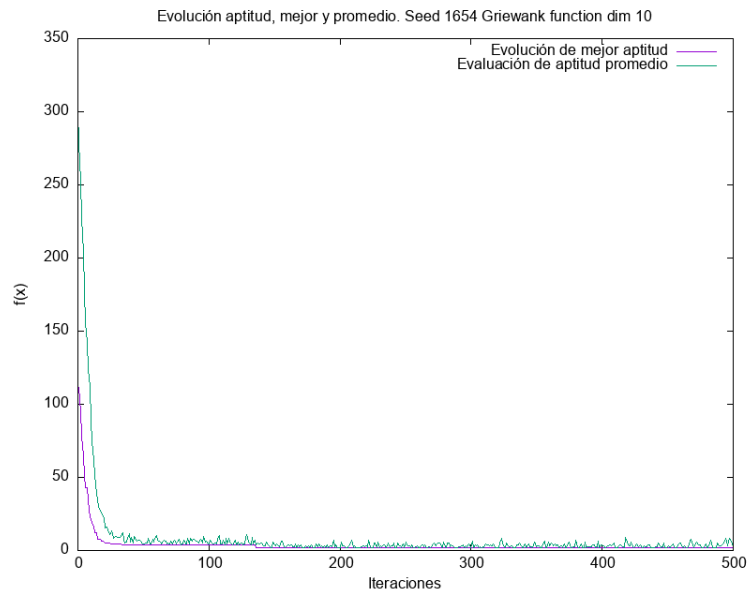
Table 5: Configuración del algoritmo de recocido simulado (SA)

Estrategia	Configuración			
	Enfriamiento	Iter modificar temp	Var modificar temp	Número de iteraciones
SA	Geométrico	1	0.999	15000
SA	Lineal	50	0.005	15000

La dimensión para el ejercicio 1 fue 10, pero luego usamos 20 y 50 para hacer una mejor comparación entre estrategias.

(b) Gráficas de evolución de aptitud para diferentes ejecuciones. **

Realizamos distintas gráficas de evolución de aptitud en dimensión 10, aquí adjuntamos un par de gráficas puesto que notamos que convergen rápido y aproximadamente a la misma velocidad, por lo tanto las gráficas son muy similares.



- (c) Gráficas de evolución promedio de aptitud. Una gráfica por cada ejemplar. En esta gráfica deberían aparecer todas los algoritmos (estrategias) que se compararon. ¿Notas alguna diferencia entre la gráfica de evolución promedio y una ejecución individual, para alguno de los algoritmos?, ¿A qué se debe? Justifica tu respuesta.

Para este inciso, utilizamos dimensión 50 para apreciar mejor el comportamiento de los algoritmos, sin embargo puede generar algunas complicaciones al comparar los resultados con la dimensión 10.

Notamos que las siguientes gráficas de los tres algoritmos genéticos tienen resultados similares, empezando ambos con el mismo valor, puesto que tienen la misma semilla para una comparación justa. Del mismo modo, los tres convergen rápidamente.

AG1 - reemplazo generacional

AG2 - reemplazo elitista

AG3 - reemplazo de peores

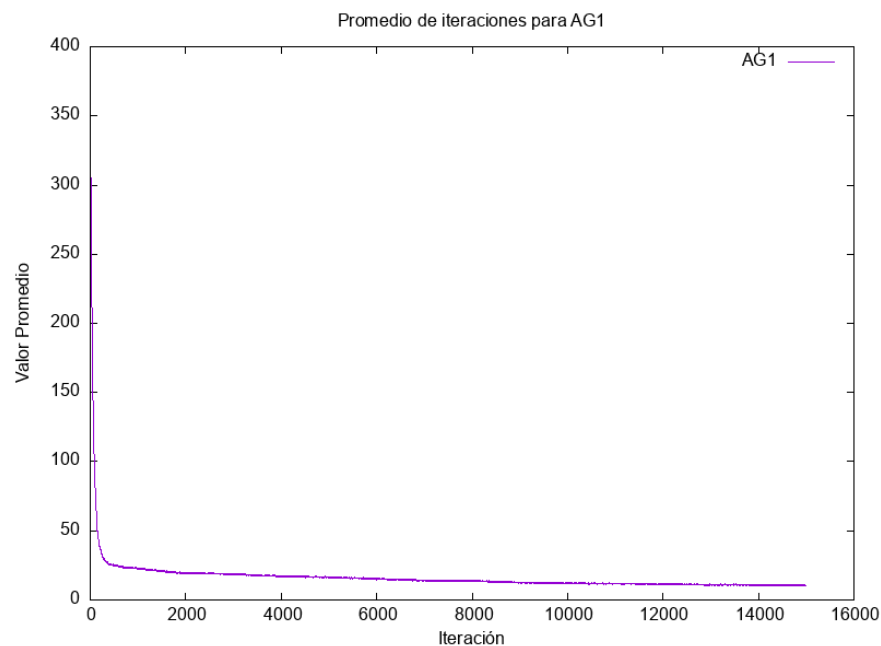
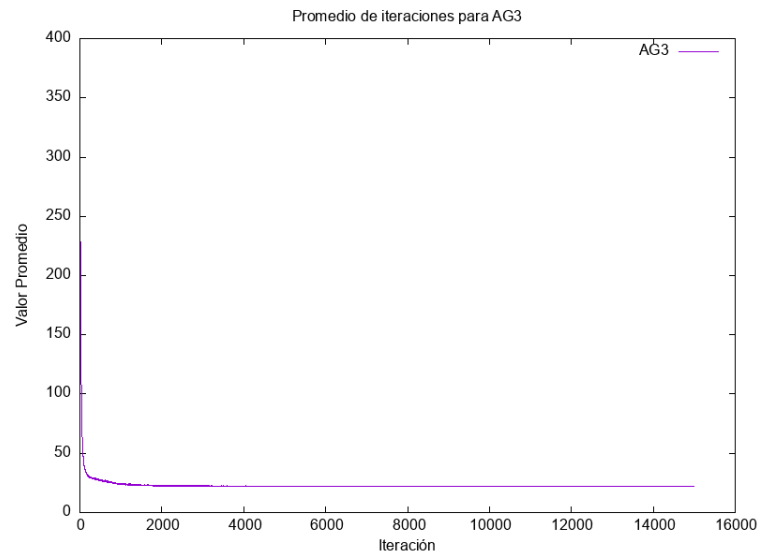
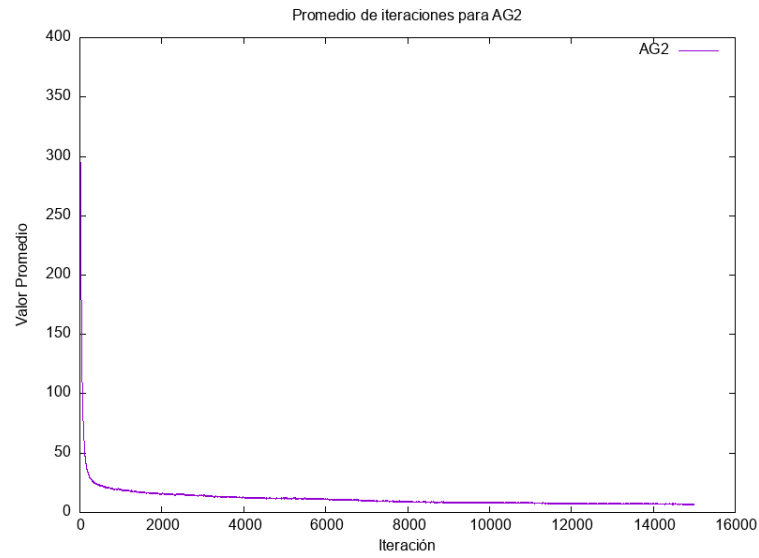


Figure 2: Gráfica AG1



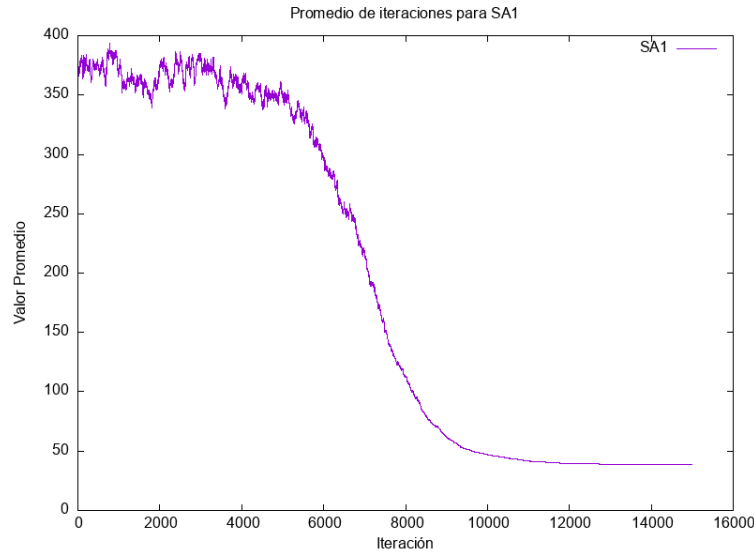


Figure 3: Gráfica SA1

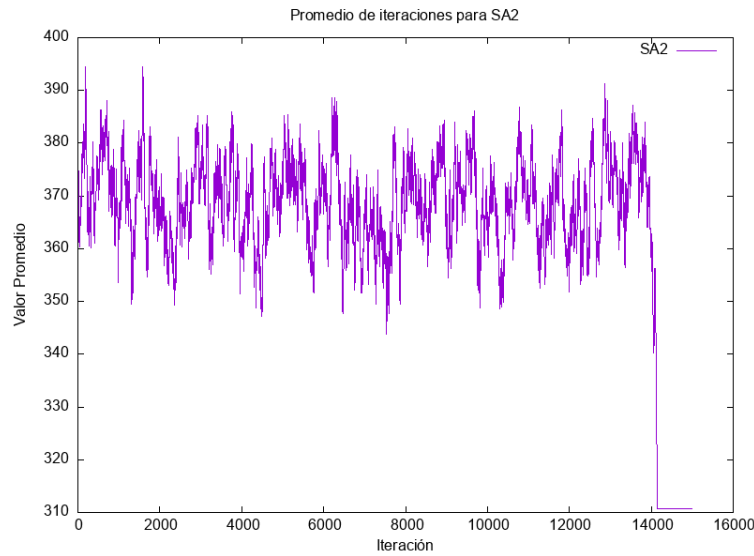


Figure 4: Gráfica SA2

De hecho esta gráfica nos indica que la implementación de este esquema de enfriamiento lineal no está funcionando como debería, parece que la temperatura no baja lo suficiente como para empezar a explotar las mejores soluciones encontradas.

Las gráficas anteriores que representan el recocido simulado de dos sistemas de enfriamiento distinto, son considerablemente diferentes, puesto que el primer sistema de enfriamiento geométrico logra acercarse al óptimo global de una forma "armónica", mientras que el segundo sistema de enfriamiento lineal, cae en un óptimo local sin posibilidad de empeorar, puesto que se enfría. Podemos concluir que la dimensión 50 es todavía demasiado para encontrar el óptimo global con nuestros algoritmos de recocido simulado.

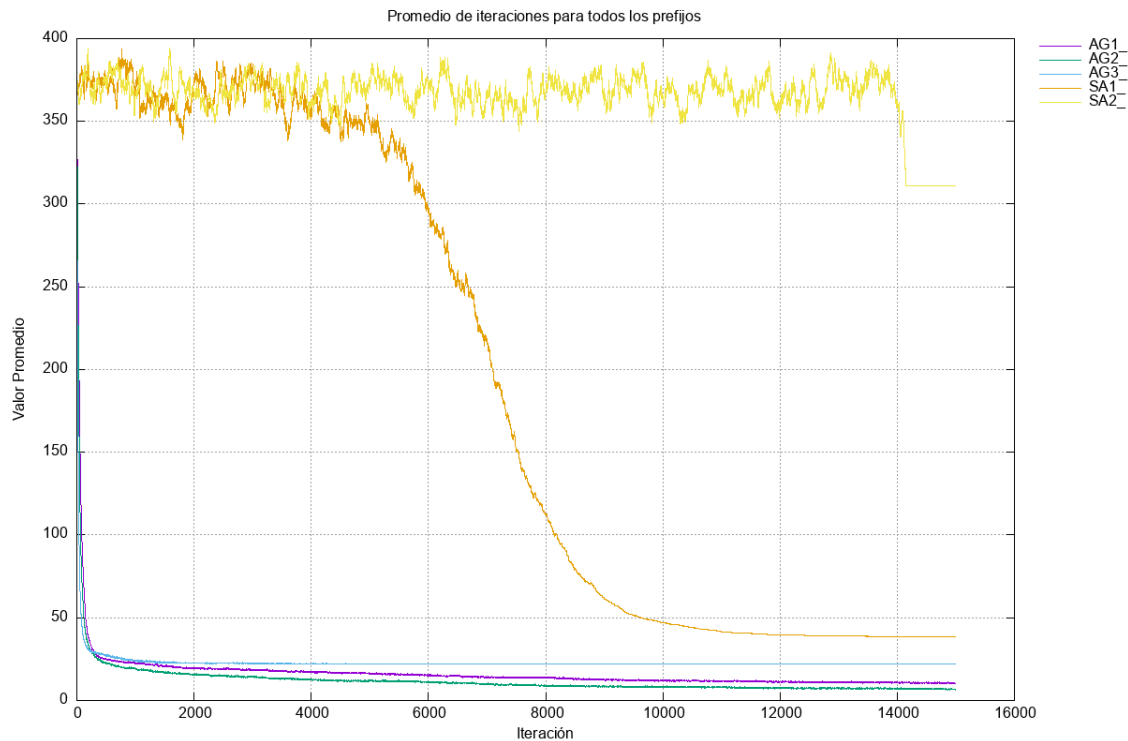


Figure 5: Comparación de todos los algoritmos anteriores

Para la comparación de todos los algoritmos de búsqueda, notamos que el mejor es el algoritmo genético con reemplazo elitista, seguido del reemplazo generacional y por último por el reemplazo de peores.

Los algoritmos de recocido simulado se quedan bastante lejos del óptimo global, puesto que la dimensión pudo haber sido un desafío para éstos tipos de enfriamiento, siendo el peor el enfriamiento lineal, el recocido simulado con enfriamiento geométrico no tuvo un desempeño tan malo, comienza de la misma forma oscilando y después continúa bajando, parece ser que a más número de iteraciones podría acercarse a los algoritmos genéticos pero ya no sería tan eficiente.

- (d) Tabla de resultados con datos estadísticos (mejor, peor, media, mediana, desviación estándar, número de iteraciones o generaciones, etc..)

Por el tiempo realizamos la tabla para un conjunto de 60 ejemplares (30 por estrategia), con las semillas mencionadas, dimensión 20 y optimizando la función Griewank.

Semillas	Estrategia	Enfriamiento	Reemplazo	Población	mejor	peor	Media	Desviación estándar	Iteraciones
[1456-1486]	SA	Geométrico	-	-	0.4476	1.0882	1.0135	0.1165	15000
	GA	-	Elitista	50	0.7991	2.4258	1.2004	0.3455	15001

Table 6: Resultados estadísticos

- (e) Gráficas de diversidad. ** Se deben implementar al menos dos medidas de distancia. Para cada medida se deben incluir las gráficas correspondientes

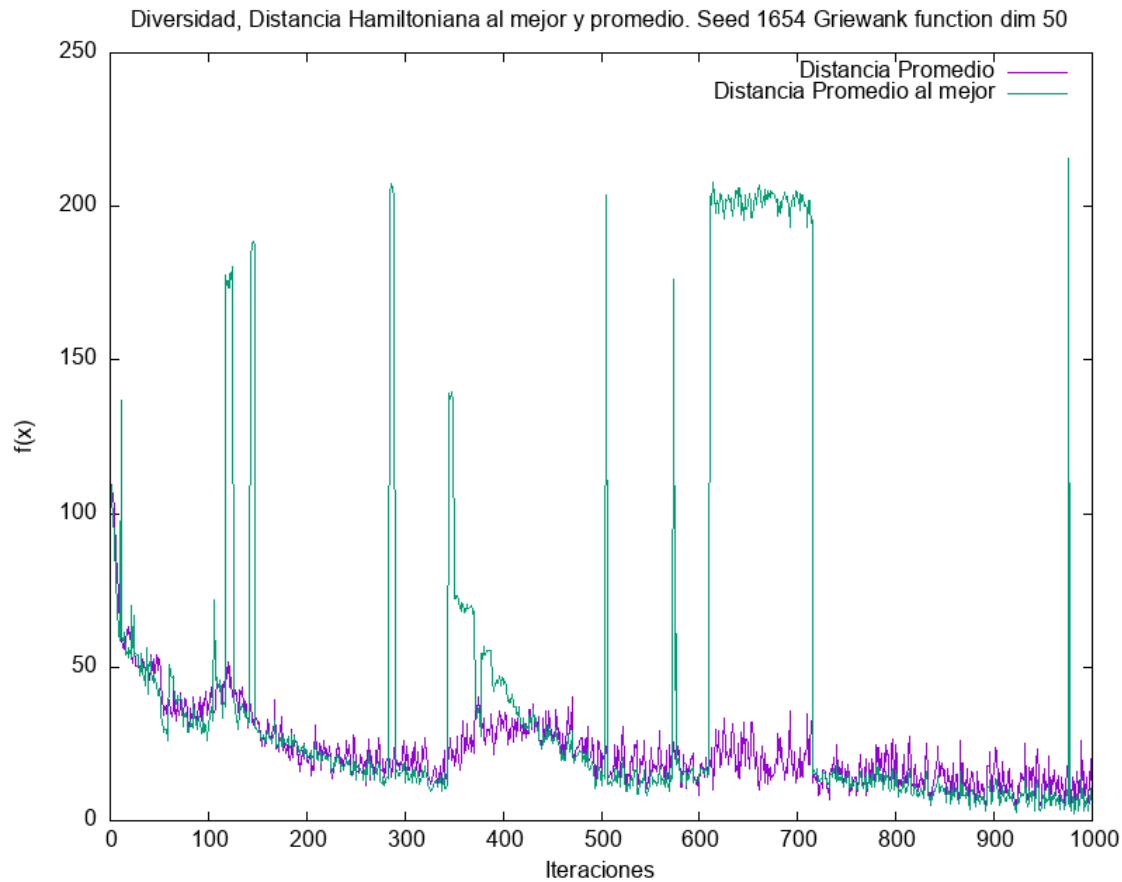


Figure 6: Distancia hamiltoniana

De esta medida de distancia podemos notar que se va acercando al cero, pero llega a tener picos muy altos y creemos que sucede al generarse por una mutación, ya que nuestro operador de mutación es la mutación flip, la cual intercambia ceros por unos.

Esto explicaría por qué la distancia crece tanto, aunque la población converge a ciertos individuos. Al hacer la mutación flip, aunque el individuo antes de ser mutado tuviera una distancia pequeña al promedio, al ser mutado la distancia ahora será muy grande, y si esa mutación hizo que ahora tenga una aptitud mejor al antiguo mejor individuo, entonces este nuevo mejor individuo será al que la demás población intente converger.

Eso explica la meseta formada entre las 600 y las 700 generaciones, ya que pudo haber aparecido un mejor que todos los demás gracias a la mutación flip, lo que también pudo causar que su distancia aumentara mucho al promedio. De esta forma en las siguientes iteraciones no habría aparecido un individuo mejor y la población estaría intentando converger hacia él.

Aunque la bajada tan rápida también nos hace pensar que apareció otro individuo con mejor aptitud pero una distancia más cercana al promedio.

Esto es una suposición de lo que pudo haber pasado.

Esto de que los individuos vayan convergiendo se ve mucho mejor en las gráficas

de frecuencia.

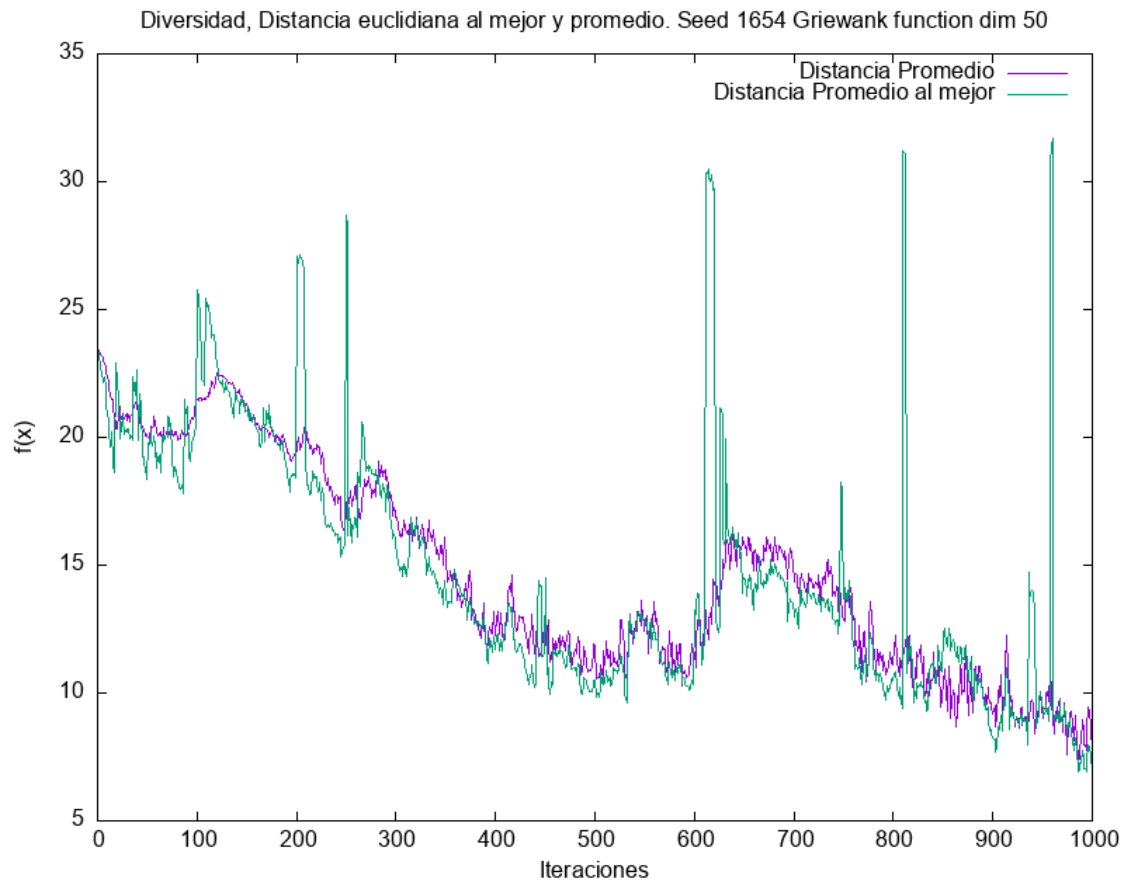


Figure 7: Distancia Euclidiana

Nuevamente entre 600 y 700 observamos subidas muy abrupta, aquí vemos que esa subida de la distancia promedio al mejor en 600 tiene una bajada más lenta. No sabemos muy bien cómo interpretar esto.

En ambas al ser medidas de distancia que cumplen la desigualdad del triángulo podemos observar que ambas se van haciendo menores mientras vamos avanzando en las iteraciones.

Esto nos dice que la diversidad de la población con el paso de las generaciones va decreciendo.

- (f) Gráficas de entropía. ** ¿Hay diferencias con respecto a las gráficas de diversidad?, ¿Por qué?

Primero para el cálculo de la entropía tuvimos que calcular las frecuencias de acuerdo a la distancia de Hamilton, al hacerlo con la población inicial obtuvimos la siguiente gráfica:

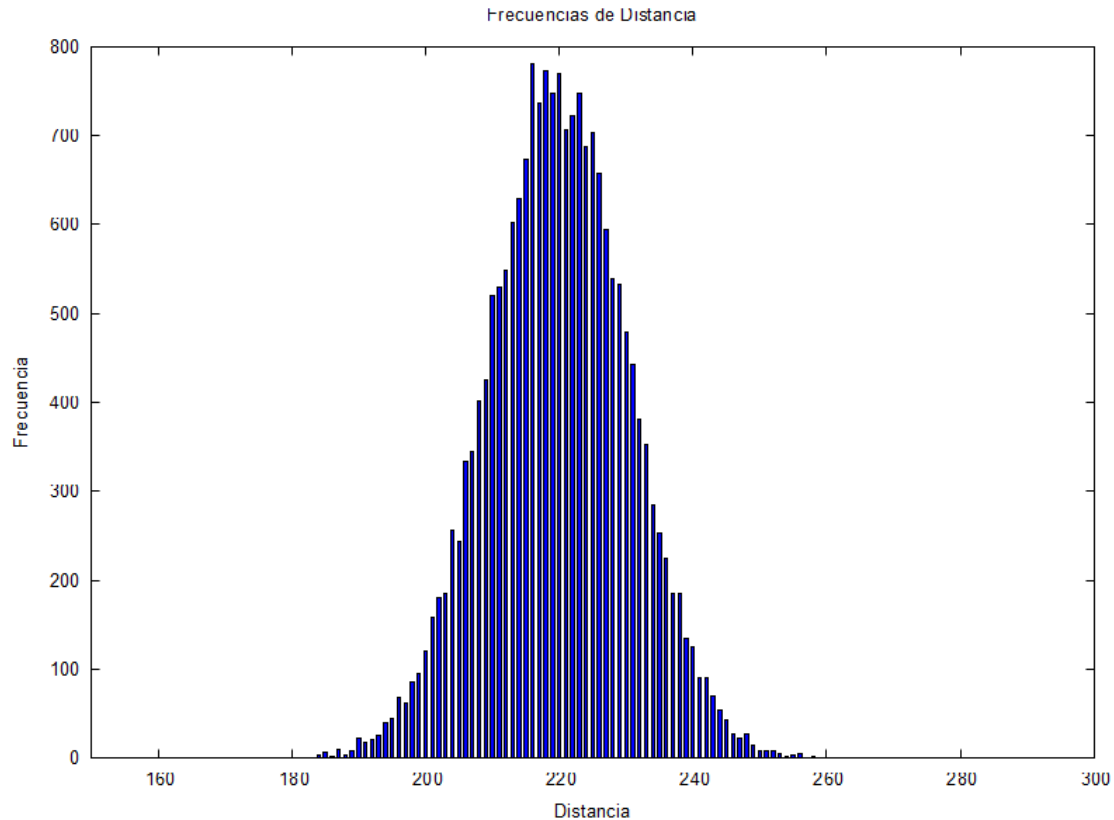


Figure 8: Distancia Euclidiana

Donde podemos notar que que las distancias se asemejan a una distribución normal, lo que era de esperar ya que la población inicial se genera de forma aleatoria, lo que haría menos probable generar individuos sean muy parecido o que sean totalmente distintos (respecto a la medida de distancia Hamiltoniana).

Por otro lado:

Al generar la gráfica de las frecuencias de la población final, la gráfica de barras aparecía aparentemente vacía. Luego al checar los valores nos dimos cuenta que todas todas las frecuencias estaban en dos distancias.

La distancia de Hamilton 0 y la distancia de Hamilton 440 eran las únicas dos con una frecuencia mayor a cero. Estos eran los extremos, es decir que para toda comparación uno a uno en individuos de la población, los individuos eran idénticos o eran totalmente distintos.

Tuvimos que separar la gráfica en dos partes para que se pudiera visualizar.

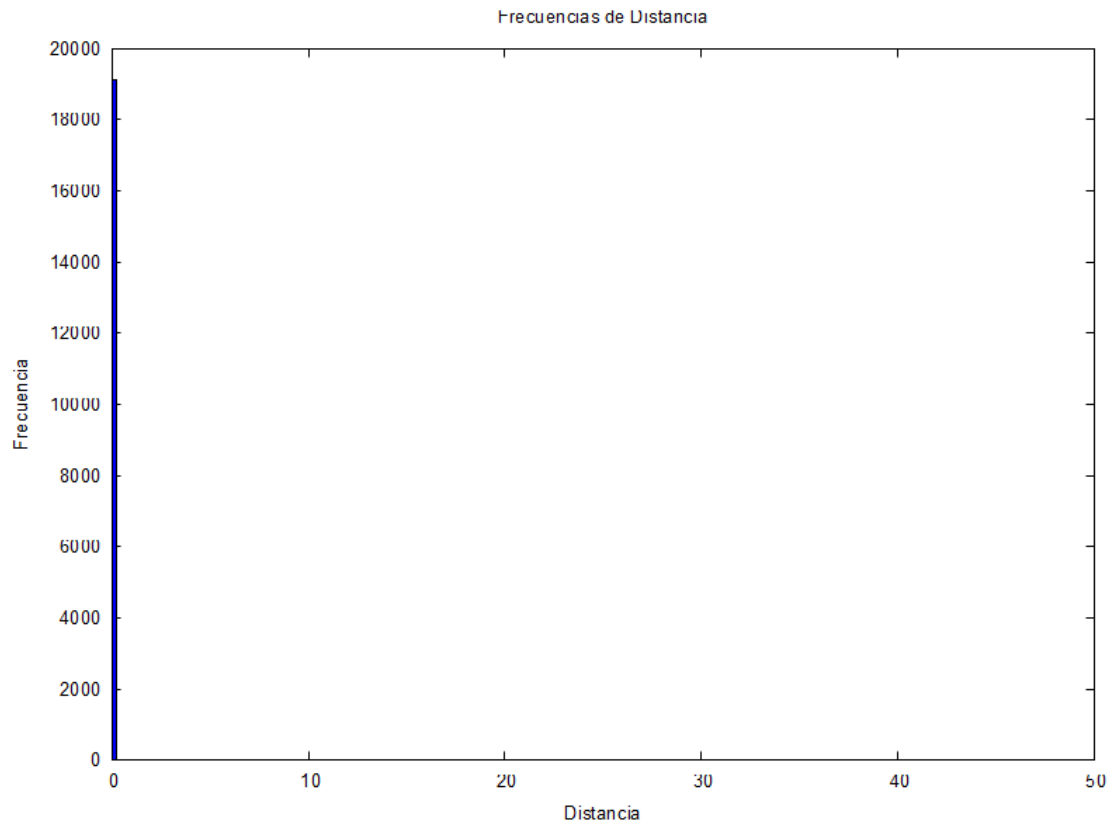


Figure 9: Frecuencia de distancia Hamilton parte 1 población final

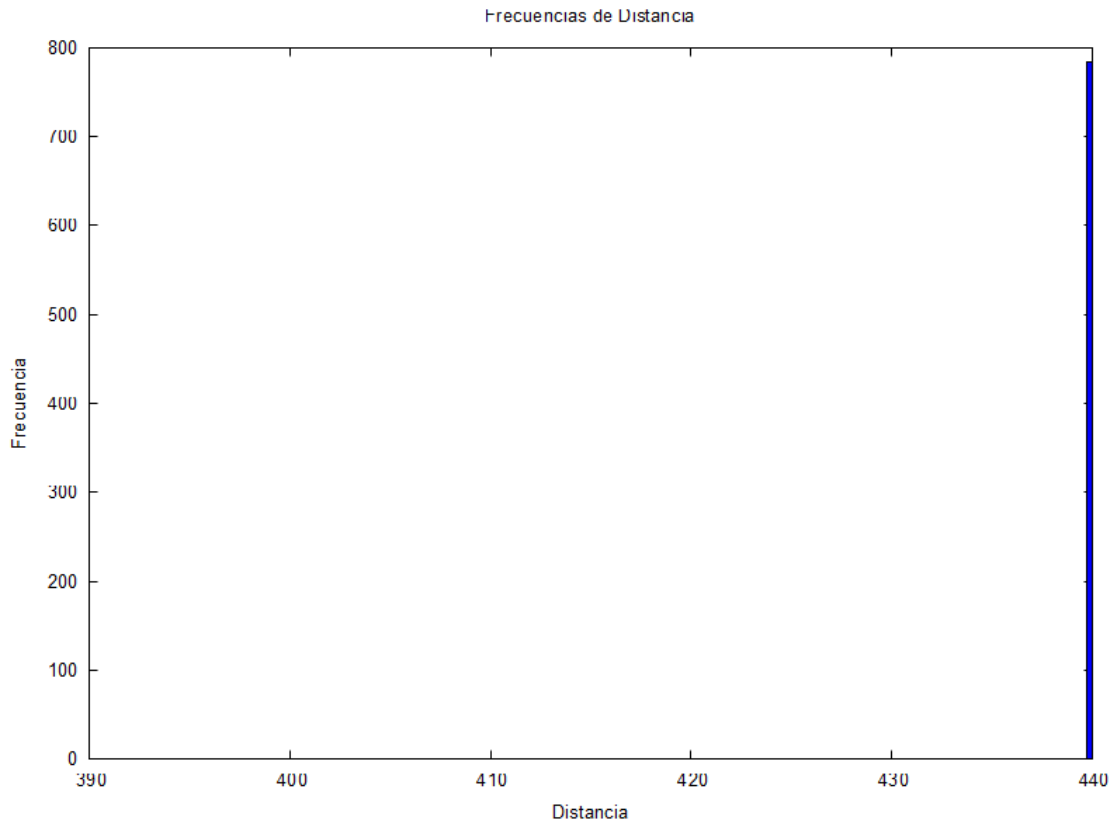


Figure 10: Frecuencia de distancia Hamilton parte 2 población final

De esta gráfica tenemos muchas suposiciones.

Creemos que la población general convergió a dos individuos; esto explicaría por que si agarras dos individuos, los individuos son idénticos o son totalmente distintos.

Es decir que la población convergió a un solo individuo y a ese individuo aplicándole la mutación flip.

Esto tiene mucho sentido porque si la población está convergiendo a un individuo con alta probabilidad de ser seleccionado por la ruleta entonces cada vez es más probable que haya duplicados de este individuo.

Al haber múltiples duplicados y estar haciendo una mutación flip a un individuo por generación, la probabilidad de que se le haga mutación flip a uno de los individuos duplicados será mayor, lo que a su vez los hará tomar un mayor trozo de la ruleta debido a que sus cuantiles se sumarían.

(g) BoxPlot.

Las gráficas de bigotes fueron generadas con dimensión 10 y los mismos parámetros de la tabla.

En la primera gráfica dejamos las 5 estrategias utilizadas.

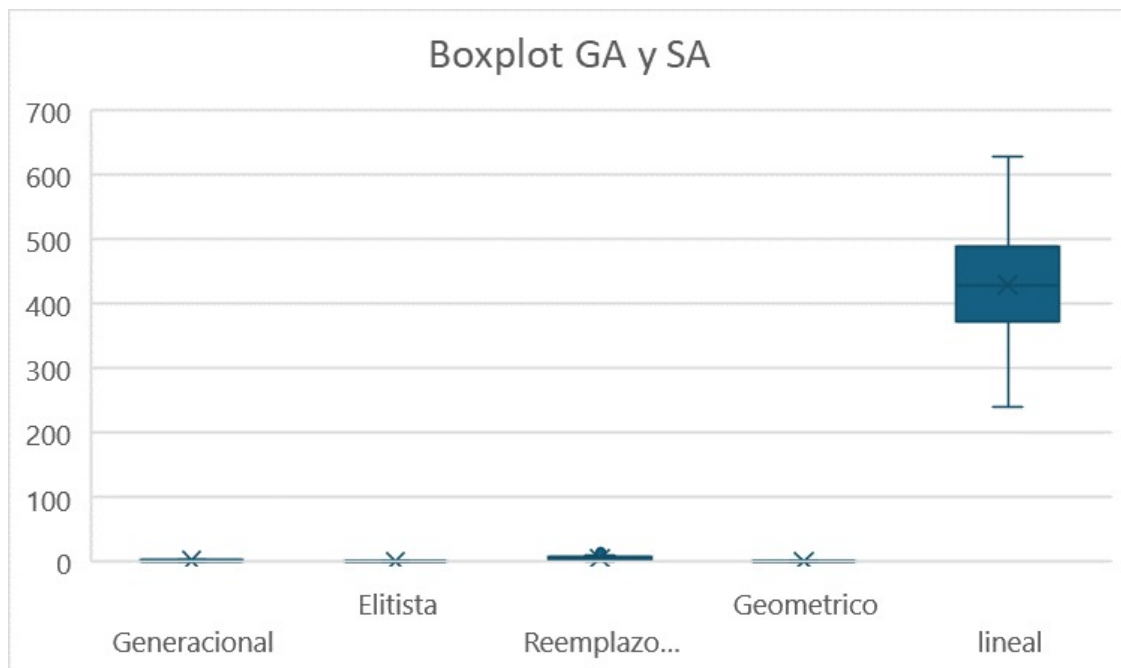


Figure 11: Distancia Euclidiana

En esta gráfica de bigotes notamos que la estrategia de recocido simulado con enfriamiento lineal fue peor que las demás estrategias parametrizadas, por lo que este estaría automáticamente descartado en una comparación para ver cuál nos sirve más para nuestras funciones.

Para observar mejor las primeras 4 estrategias parametrizadas graficamos una nueva boxplot.

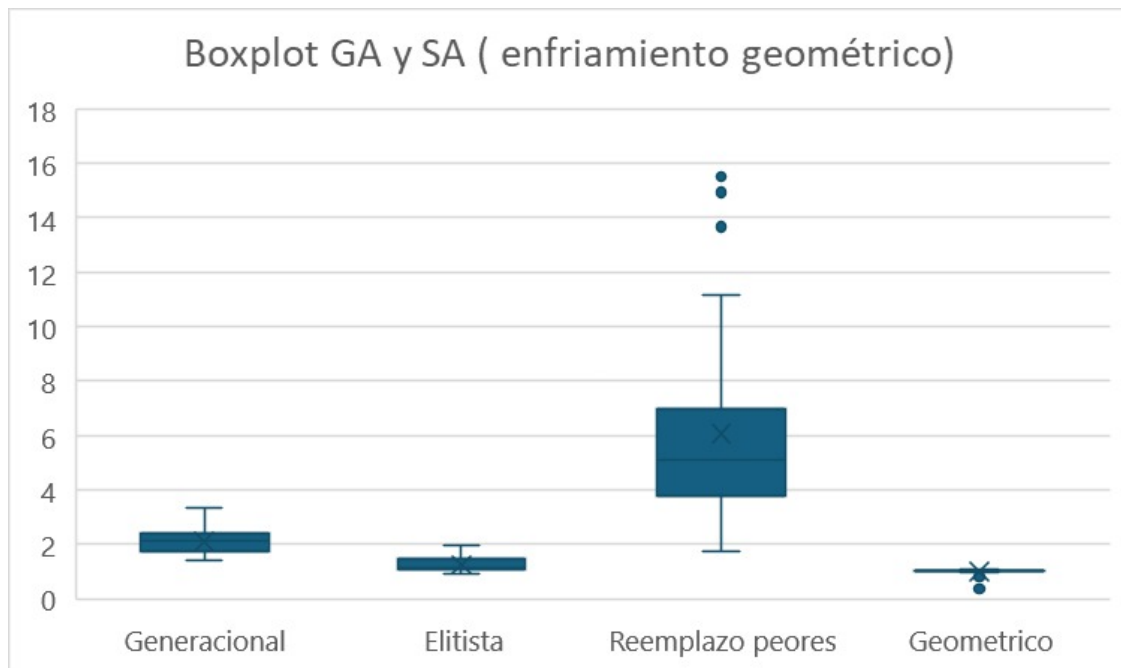


Figure 12: Distancia Euclidiana

En esta gráfica de bigotes notamos que el recocido simulado con el enfriamiento geométrico nos asegura una mejor solución que las demás estrategias y en las otras estrategias aunque podemos obtener una solución similar, también corremos el riesgo de obtener una solución peor a la del geométrico.

Por lo que para la dimensión 10, nos conviene más usar el recocido simulado con enfriamiento geométrico.