

Memoria Parte 1

Mikel García

Table of Contents

1	Introducción.....	2
2	Implementación.....	2
3	Algoritmos Genéticos.....	2
3.1	Reporte Individual.....	2
3.1.1	Ejecución.....	2
3.1.1.1	GSA.....	3
3.1.1.2	PSO.....	4
3.1.1.3	GA.....	5
3.1.2	Estadísticos.....	6
3.1.2.1	GSA.....	6
3.1.2.2	PSO.....	7
3.1.2.3	GA.....	7
3.1.3	Velocidad de convergencia.....	7
3.1.3.1	GSA.....	7
3.1.3.2	PSO.....	8
3.1.3.3	GA.....	9
3.1.4	Efecto del tamaño de la población.....	10
3.1.4.1	GSA.....	10
3.1.4.2	PSO.....	11
3.1.4.3	GA.....	12
3.1.5	Estabilidad estadística.....	13
3.1.5.1	GSA.....	13
3.1.5.2	PSO.....	14
3.1.5.3	GA.....	15
3.2	Comparación.....	16
4	Modelos Surrogados.....	18
4.1	Efecto del tamaño de la muestra.....	18
4.2	Error del surrogado.....	21

1 Introducción

En esta memoria se recogen los resultados de los tres algoritmos genéticos vistos en clase: una implementación del Algoritmo Genético (GA), una implementación del Algoritmo del enjambre (PSO) y una implementación del Algoritmo de Búsqueda Gravitacional (GSA). Se estudia la velocidad de convergencia del algoritmo, así como el efecto del tamaño de la población en el algoritmo o el efecto de las condiciones estocásticas. También, se estudia el efecto del uso de un modelo surrogado como sustituto de la función fitness.

2 Implementación

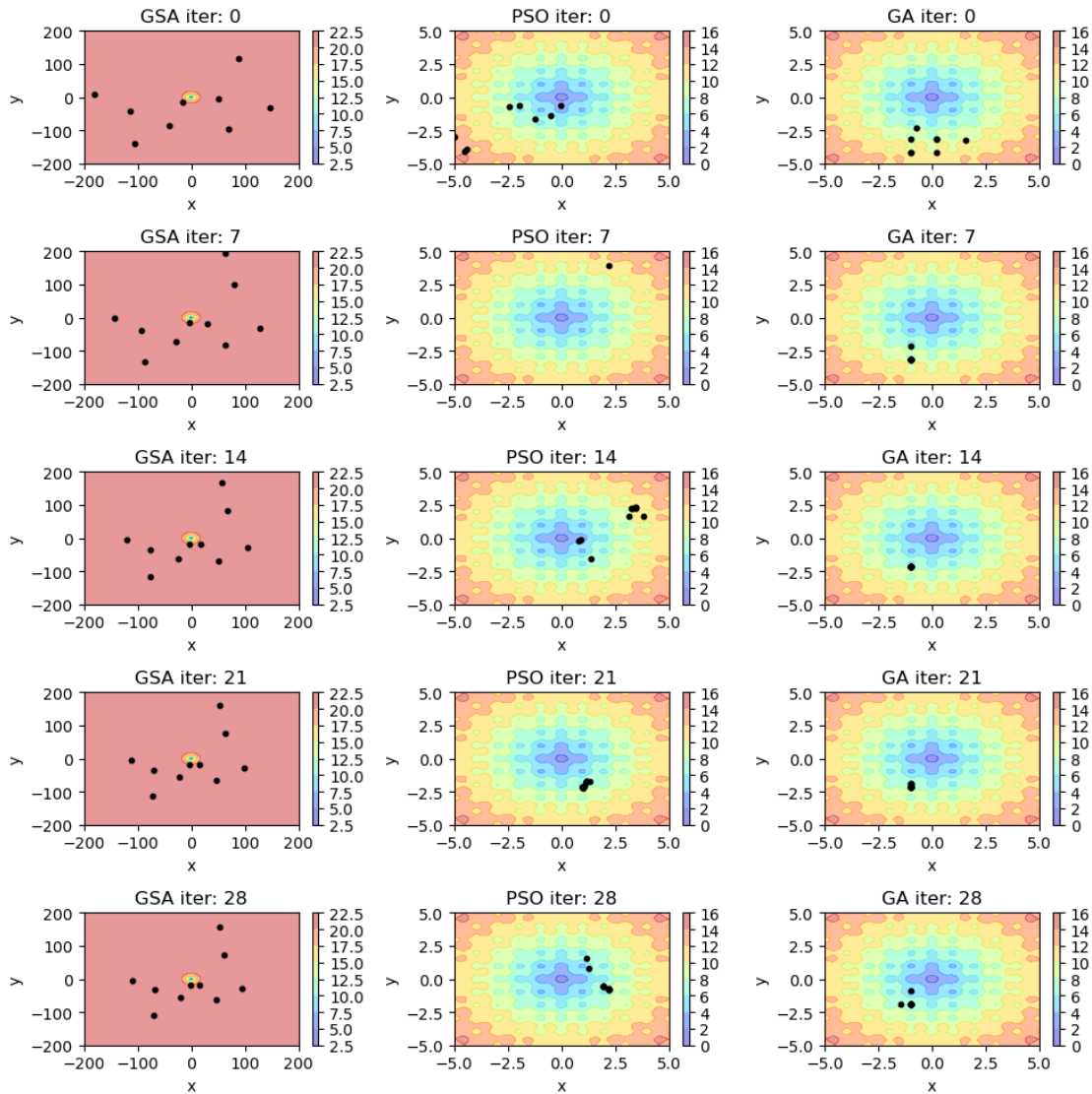
Para todos los algoritmos se han implementado una serie de operadores para evolucionar la población. Como mejora, se ha introducido un modelo surrogado en todos ellos que se empieza a utilizar al registrarse suficientes individuos para formar una muestra. Se registran varios estadísticos con cada iteración de la población, hay tres colecciones de estadísticos; para la función fitness, para el modelo surrogado y para el error relativo que se comete al utilizar el modelo surrogado. El tamaño de la población por defecto es 10 de individuos y el valor de la semilla es 199 salvo que se indique lo contrario.

3 Algoritmos Genéticos

3.1 Reporte Individual

3.1.1 Ejecución

Tomando las primeras ejecuciones de los algoritmos para la función ackley se observa como conforme se actualiza la población las partículas se mueven en las dos dimensiones del espacio por el efecto que tienen unas sobre otras. GSA posee gran variabilidad en cuanto a la distribución de las partículas. Por otro lado, PSO y GA se encuentran en intervalos más pequeños para las dos componentes de las partículas. PSO parece mostrar más variabilidad a lo largo de las iteraciones moviéndose de una región de la función a otra pese a encontrarse cerca del óptimo global. GA, por otra parte, inicia con menos variabilidad que los otros dos algoritmos y al cabo de ciertas iteraciones la mayoría de los puntos se encuentran en un mismo óptimo. Es en este momento que los operadores de cruce y mutación permiten que el algoritmo avance.

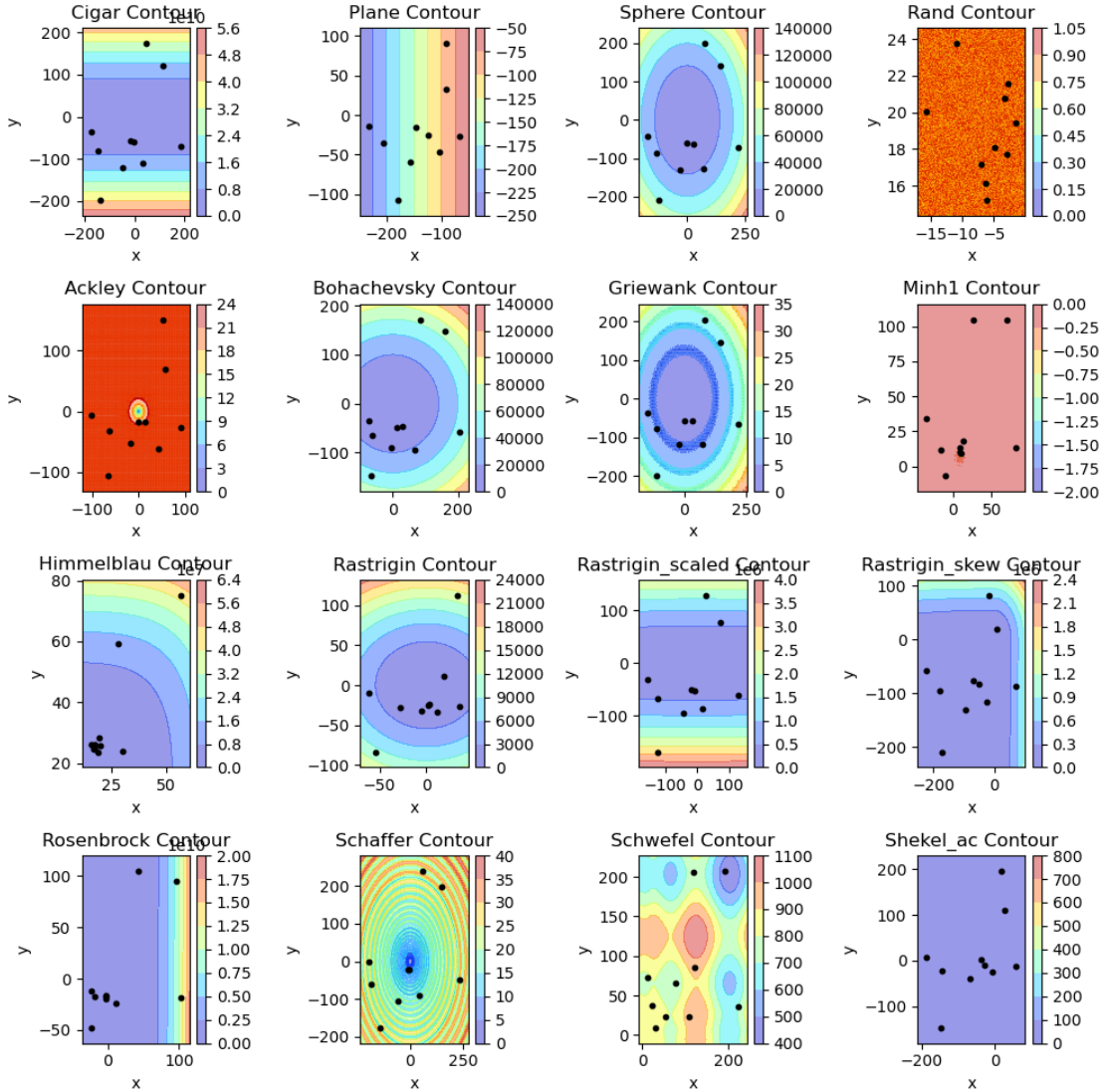


Gráfica 1: Evolución de las partículas a lo largo de varias iteraciones

Podemos comprobar el resultado de los algoritmos para una población de 10 individuos al cabo de 100 iteraciones.

3.1.1.1 GSA

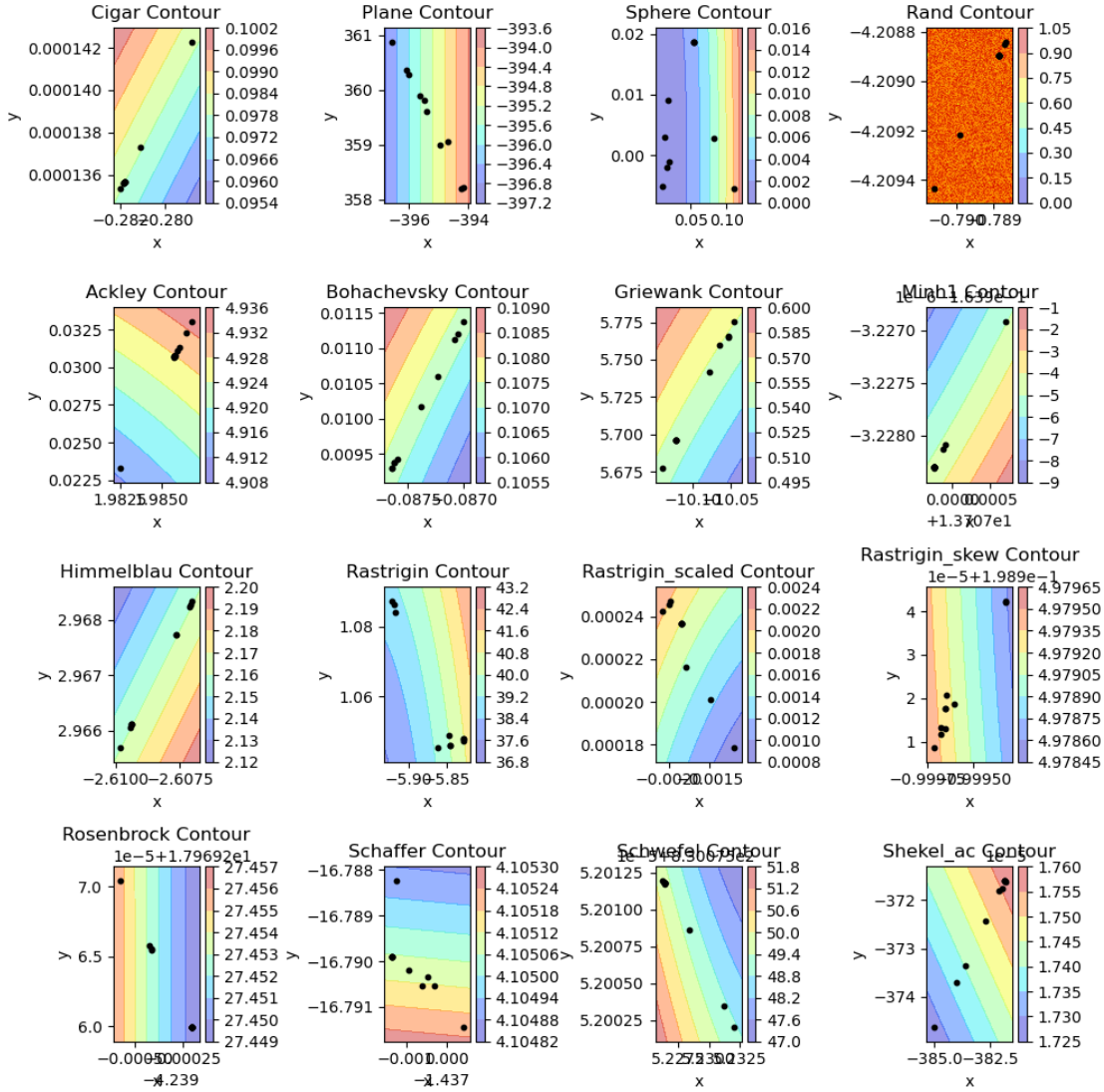
GSA no parece tener problemas para hallar óptimos próximos al óptimo global siempre y cuando la función cuente con cierta monotonía. Si bien, la escala de las componentes muestra valores muy grandes que implican una alta dispersión entre partículas.



Gráfica 2: configuración final de las partículas con el algoritmo GSA

3.1.1.2 PSO

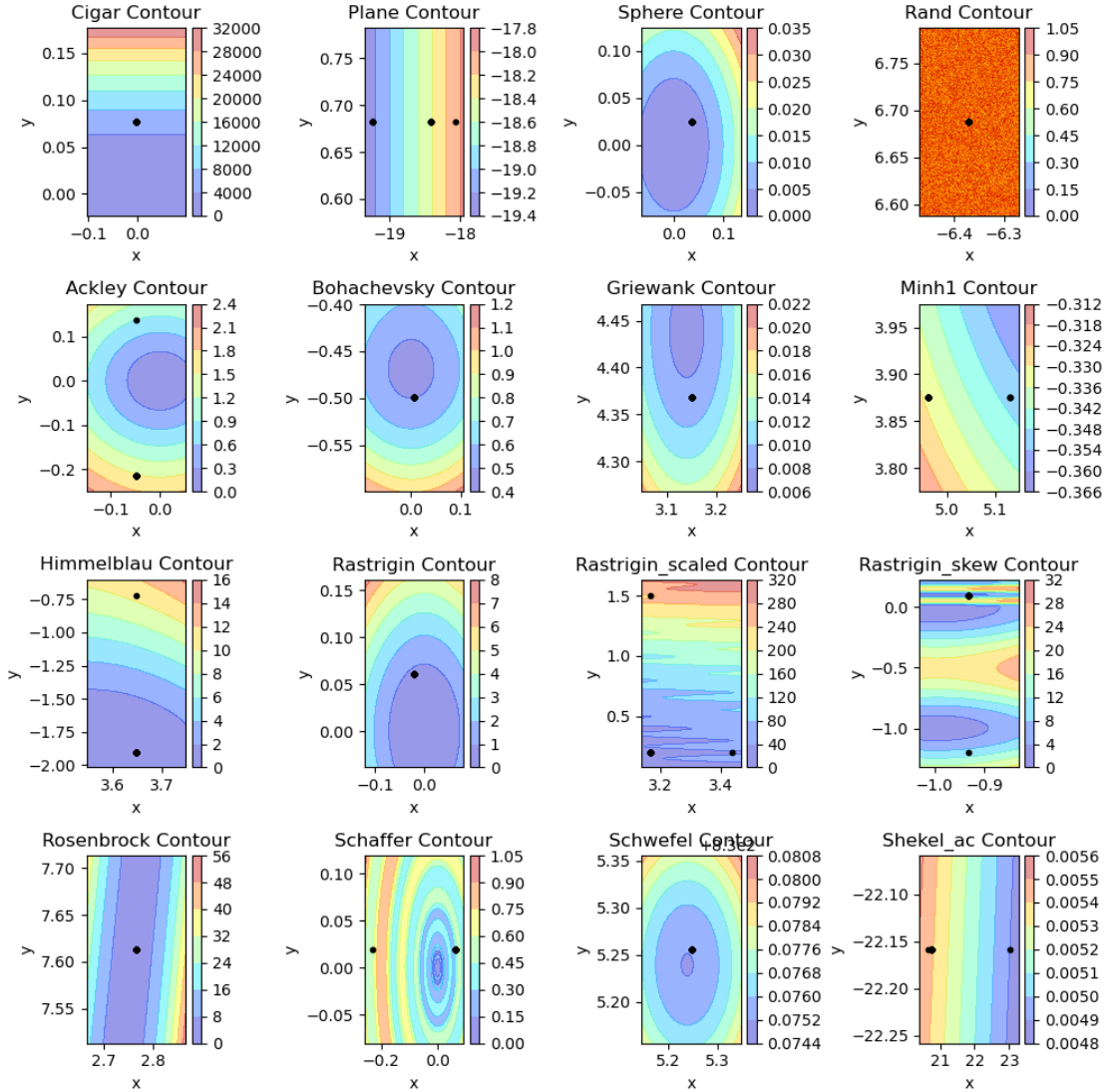
En general, el algoritmo parece estancarse en regiones intermedias. Se puede apreciar que las partículas están muy próximas al observar el valor de la escala.



Gráfica 3: configuración final de las partículas con el algoritmo PSO

3.1.1.3 GA

El algoritmo converge a un único punto en la mayoría de los casos. Queda cerca del óptimo global pero no llega a alcanzarlo.



Gráfica 4: configuración final de las partículas con el algoritmo GA

3.1.2 Estadísticos

En las siguientes tablas se muestran los resultados promedio de los tres algoritmos de optimización para las distintas generaciones y estadísticos de interés. Nótese, solo se muestra una de cada 20 iteraciones para que las tablas no sean excesivamente largas. Igualmente, las tablas contienen el valor de los estadísticos para la función objetivo rand. El general el comportamiento de los algoritmos es muy similar, la media, aunque parece no converger a ningún valor parece estabilizarse entorno a 0.4, mientras que la desviación se encuentra entorno a 0.2 y 0.3. Respecto al mínimo y al máximo, siempre toman valores cercanos a 0 y a 1 respectivamente pero nunca iguales.

3.1.2.1 GSA

Generación	Avg	Std	Min	Max	Evals
20	0.536031	0.288717	0.040758	0.928239	10

Generación	Avg	Std	Min	Max	Evals
40	0.396331	0.241362	0.015232	0.871844	10
60	0.440099	0.264747	0.043275	0.839668	10
80	0.582197	0.313709	0.079227	0.977772	10
100	0.526638	0.333109	0.030873	0.953275	10

Tabla 1: Valores de los estadísticos en GSA de la función rand

3.1.2.2 PSO

Generación	Avg	Std	Min	Max	Evals
20	0.436475	0.321481	0.061497	0.923702	10
40	0.468950	0.194351	0.074661	0.747077	10
60	0.537302	0.342462	0.048334	0.998624	10
80	0.386911	0.268541	0.004217	0.929089	10
100	0.535587	0.263124	0.017657	0.970615	10

Tabla 2: Valores de los estadísticos en PSO de la función rand

3.1.2.3 GA

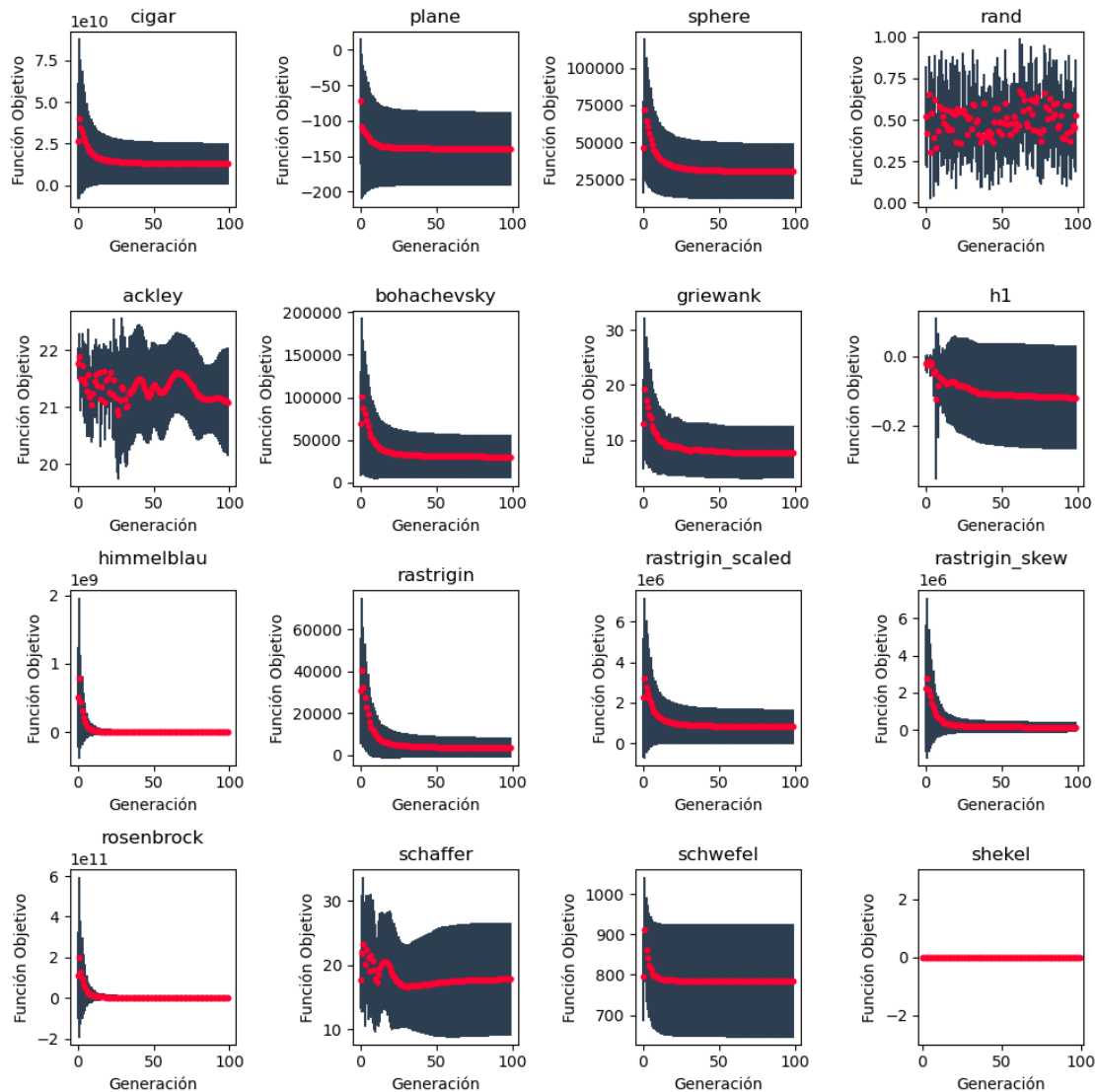
Generación	Avg	Std	Min	Max	Evals
20	0.555402	0.377241	0.102673	0.985795	10
40	0.566746	0.286752	0.136741	0.979465	10
60	0.476381	0.330814	0.002333	0.896908	10
80	0.377818	0.334051	0.051377	0.996816	10
100	0.433664	0.314667	0.073394	0.960589	10

Tabla 3: Valores de los estadísticos en GA de la función rand

3.1.3 Velocidad de convergencia

3.1.3.1 GSA

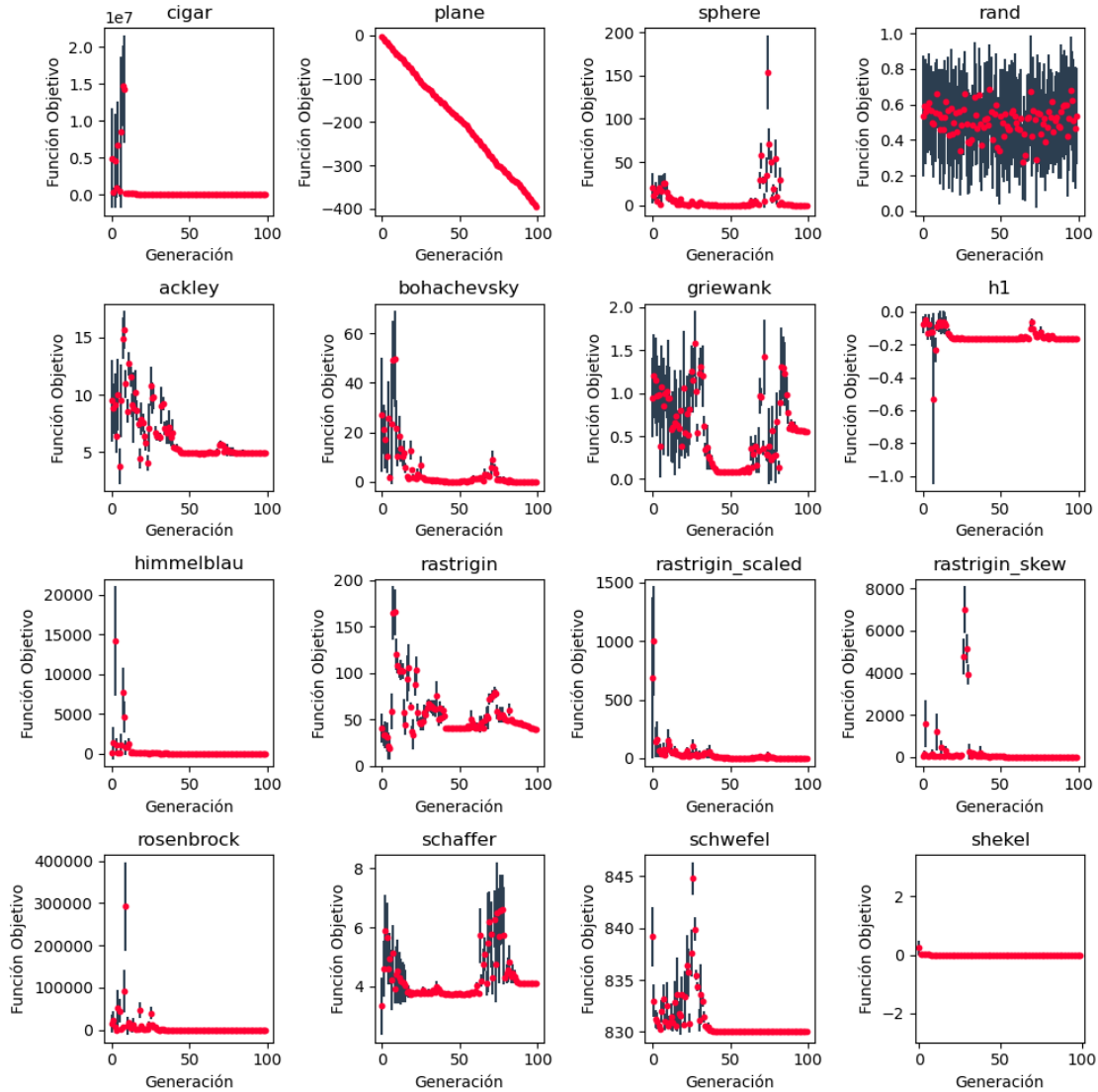
La velocidad de convergencia depende en gran medida de la función objetivo. En general GSA halla una configuración estable al cabo de pocas iteraciones. Si bien, hay algunas funciones, como rand, ackley, h1 y schaffer para las que GSA tarda más de 20 iteraciones en llegar a una configuración estable o no llega a encontrar dicha configuración. En general parece exhibir gran variabilidad entre las distintas instancias dentro de una misma iteración, aunque también hay casos, como la función himmelblau, rosenbrock o shekel donde la variabilidad entre las partículas es poco apreciable. Esto sugiere dos tipos de comportamiento en el algoritmo; en primer lugar, si hay alta variabilidad implica que las partículas han quedado atrapadas en óptimos locales distintos, en segundo lugar, cuando apenas hay variabilidad las partículas quedan agrupadas en torno a un mismo óptimo o una ravine.



Gráfica 5: Media y desviación de las funciones fitness de GSA

3.1.3.2 PSO

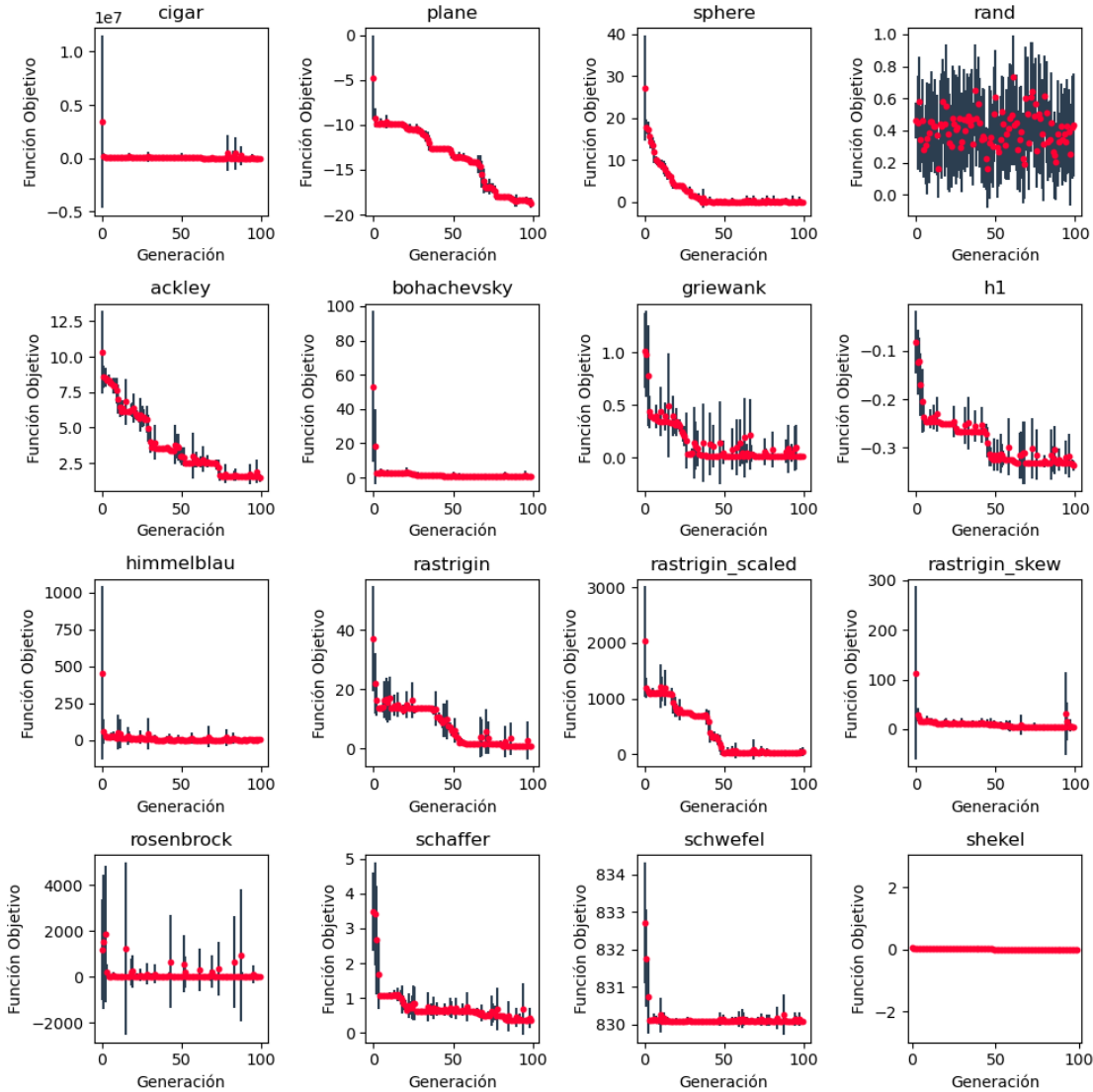
En el caso de PSO para la mayoría de las funciones el valor medio de la función objetivo disminuye con el tiempo y, en algunos casos, la variabilidad aumenta sin previo aviso. Este comportamiento sugiere que, aunque el óptimo global sea el mismo para todas las partículas, el último óptimo de la partícula se ha actualizado de manera que globalmente aumenta la dispersión entre partículas. Es el caso de la función griewank, rastrigin, schaffer o schwefel.



Gráfica 6: Media y desviación de las funciones fitness de PSO

3.1.3.3 GA

En el caso de GA, en general converge a un valor con pocas iteraciones. También ocurre que, una vez se estabiliza en un valor, hay un aumento de variabilidad que permite hallar mejores valores de la función objetivo.



Gráfica 7: Media y desviación de las funciones fitness de GA

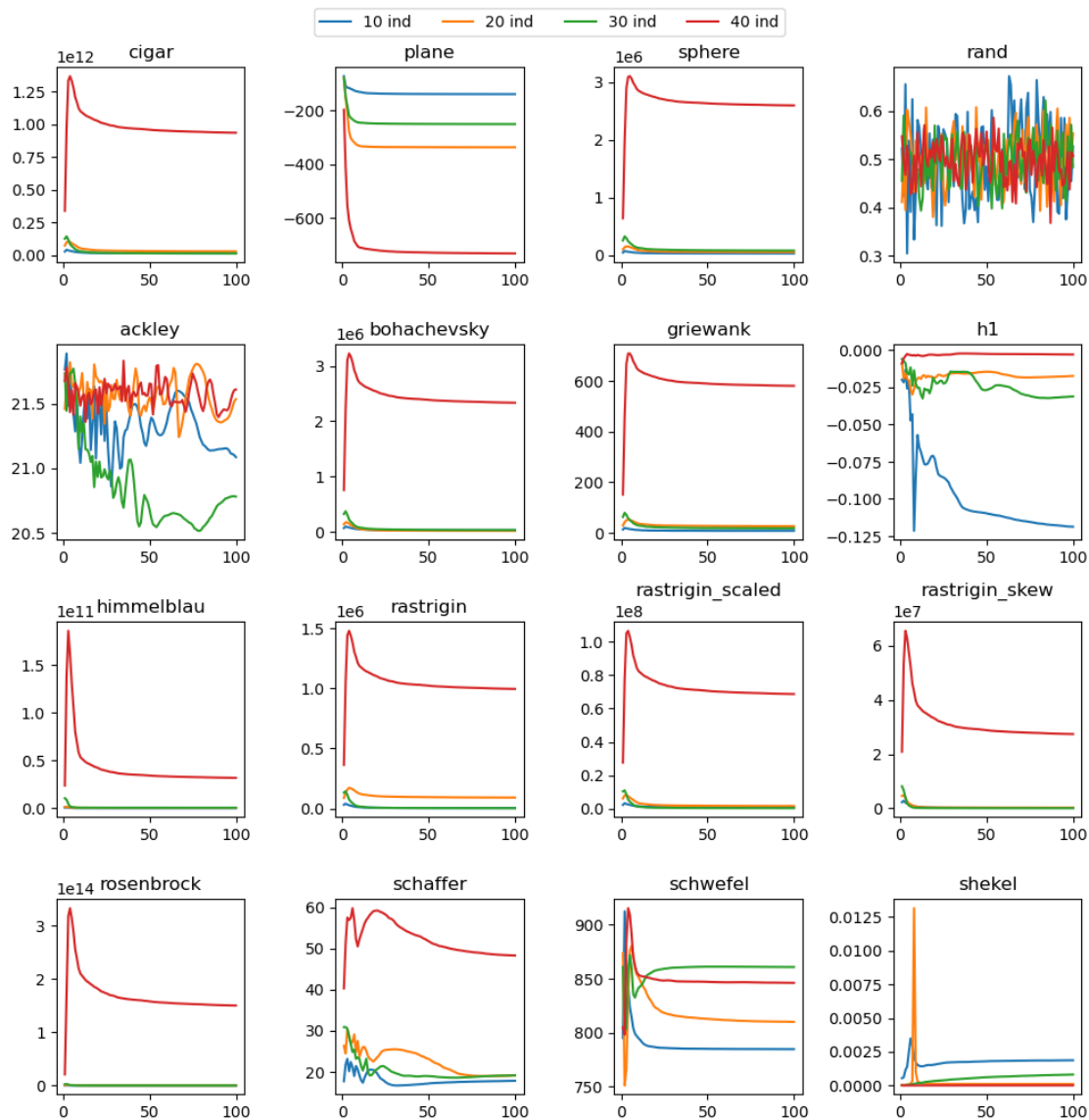
3.1.4 Efecto del tamaño de la población

En líneas generales se ve que hay algunas funciones que se benefician del tamaño de la población para hallar un óptimo. A excepción de la función rand, donde el efecto de la población se aprecia en la variabilidad de estadístico, en lugar de en los valores que toma conforme se actualiza la población.

3.1.4.1 GSA

En términos generales el algoritmo parece beneficiarse más de tamaños de población pequeños frente a grandes. Hay una excepción, que es el caso de la función plane, donde claramente un tamaño de población más grande ayuda a encontrar un óptimo mejor en el mismo número de iteraciones. Es muy posible que dada la dispersión entre las partículas un tamaño más pequeño propicie que las partículas se acaben

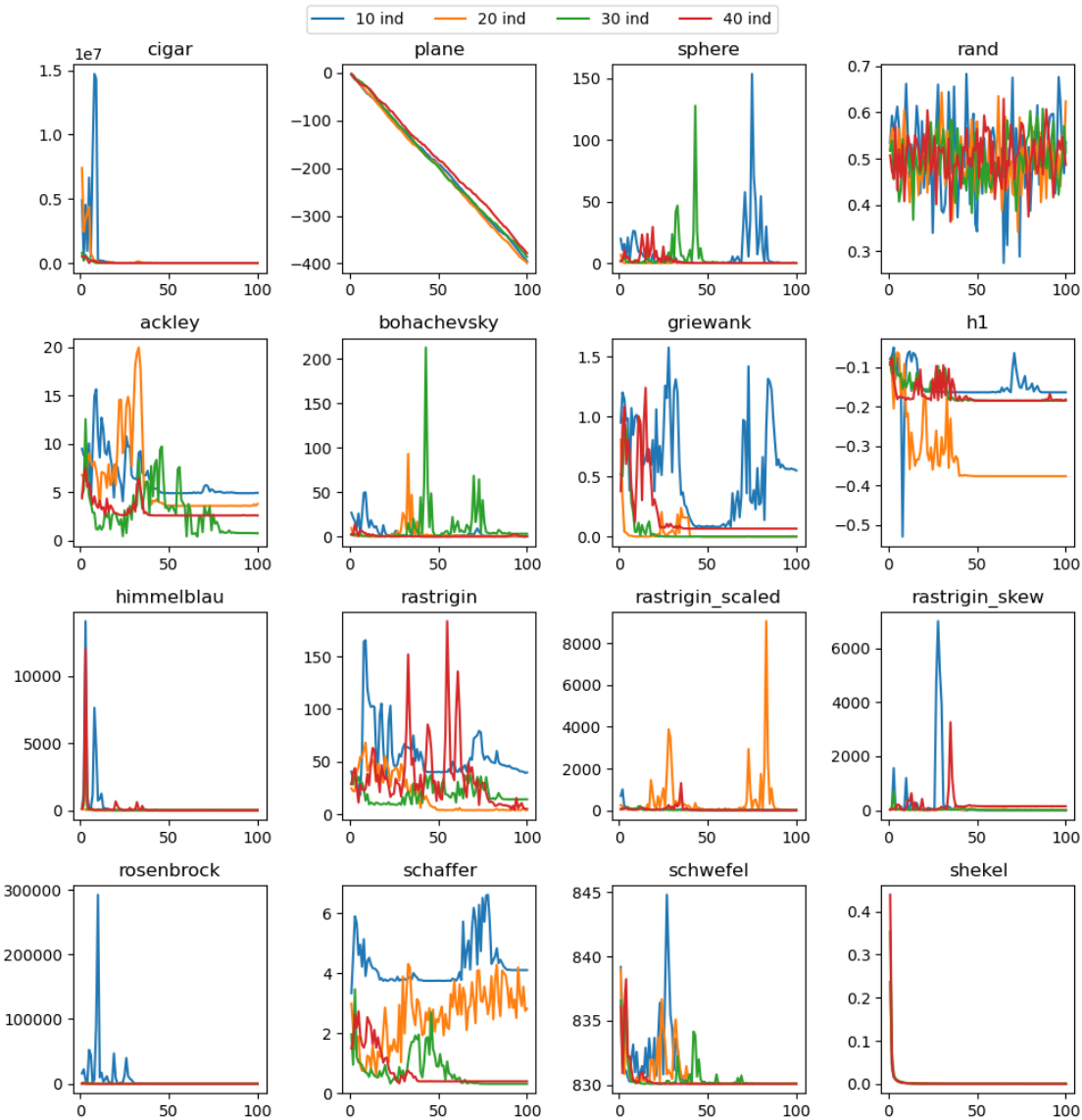
atrayendo a un mismo punto con más rapidez. Mientras que en tamaños más grandes se da esta misma convergencia, pero en distintos óptimos, de manera que puede haber un gran grupo de partículas que estén sesgando el valor de la función objetivo.



Gráfica 8: Valor medio de la función objetivo según el tamaño de la población para GSA

3.1.4.2 PSO

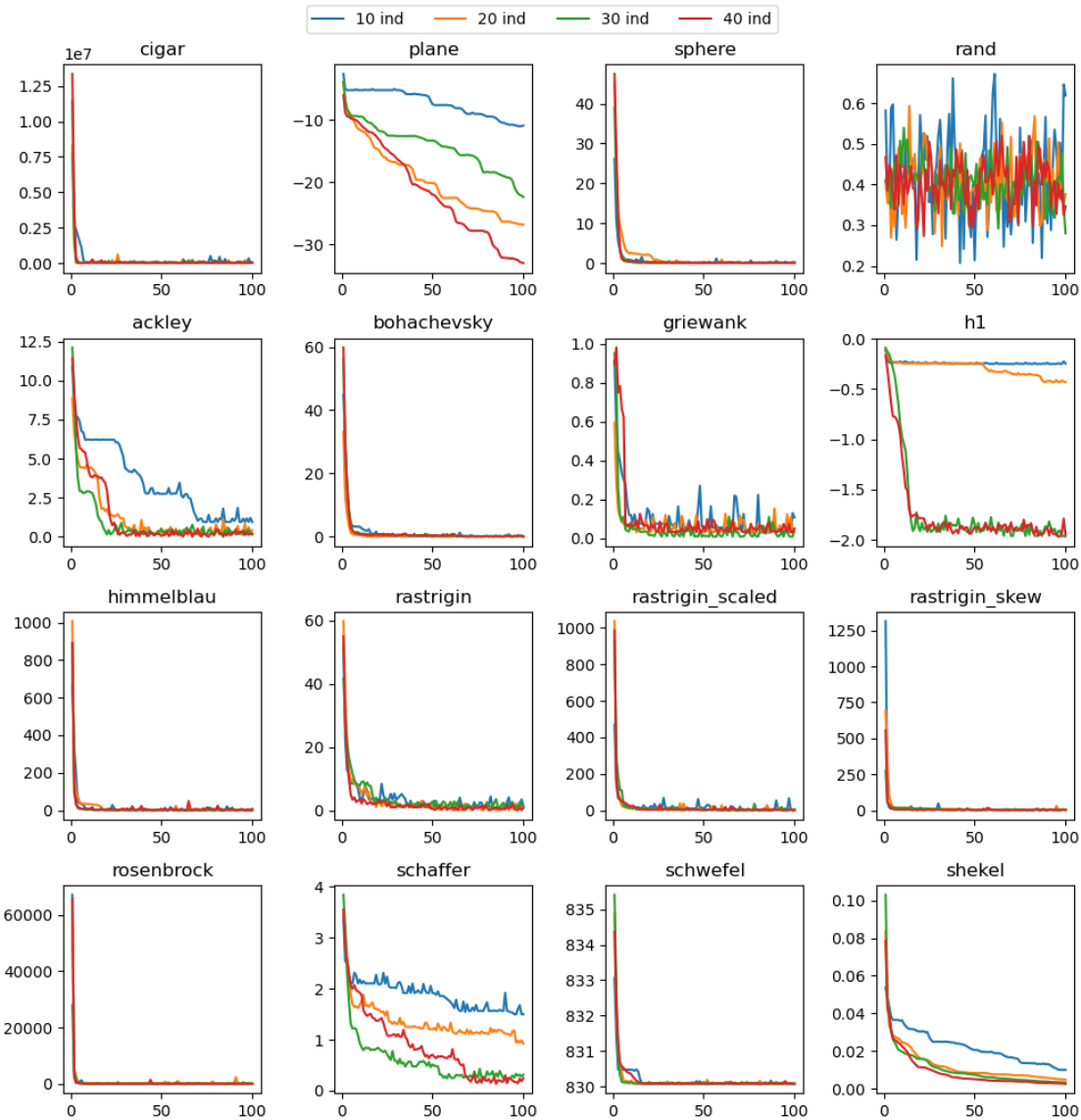
Hay pocas funciones para las que el algoritmo no muestre un comportamiento volátil a lo largo de la ejecución. No es del todo claro si es preferible utilizar un tamaño de población pequeño frente a uno grande, aunque parece que en general a mayor tamaño poblacional es esperable un menor valor medio de la función objetivo.



Gráfica 9: Valor medio de la función objetivo según el tamaño de la población para PSO

3.1.4.3 GA

Para este algoritmo es claro que se obtienen valores de la función objetivo más bajos a la larga a más individuos tiene la población. Dado que tenemos dos operadores probabilísticos, mutación y cruce, a mayor tamaño poblacional más probabilidades de que estos operadores generen una configuración mejor de la solución.



Gráfica 10: Valor medio de la función objetivo según el tamaño de la población para GA

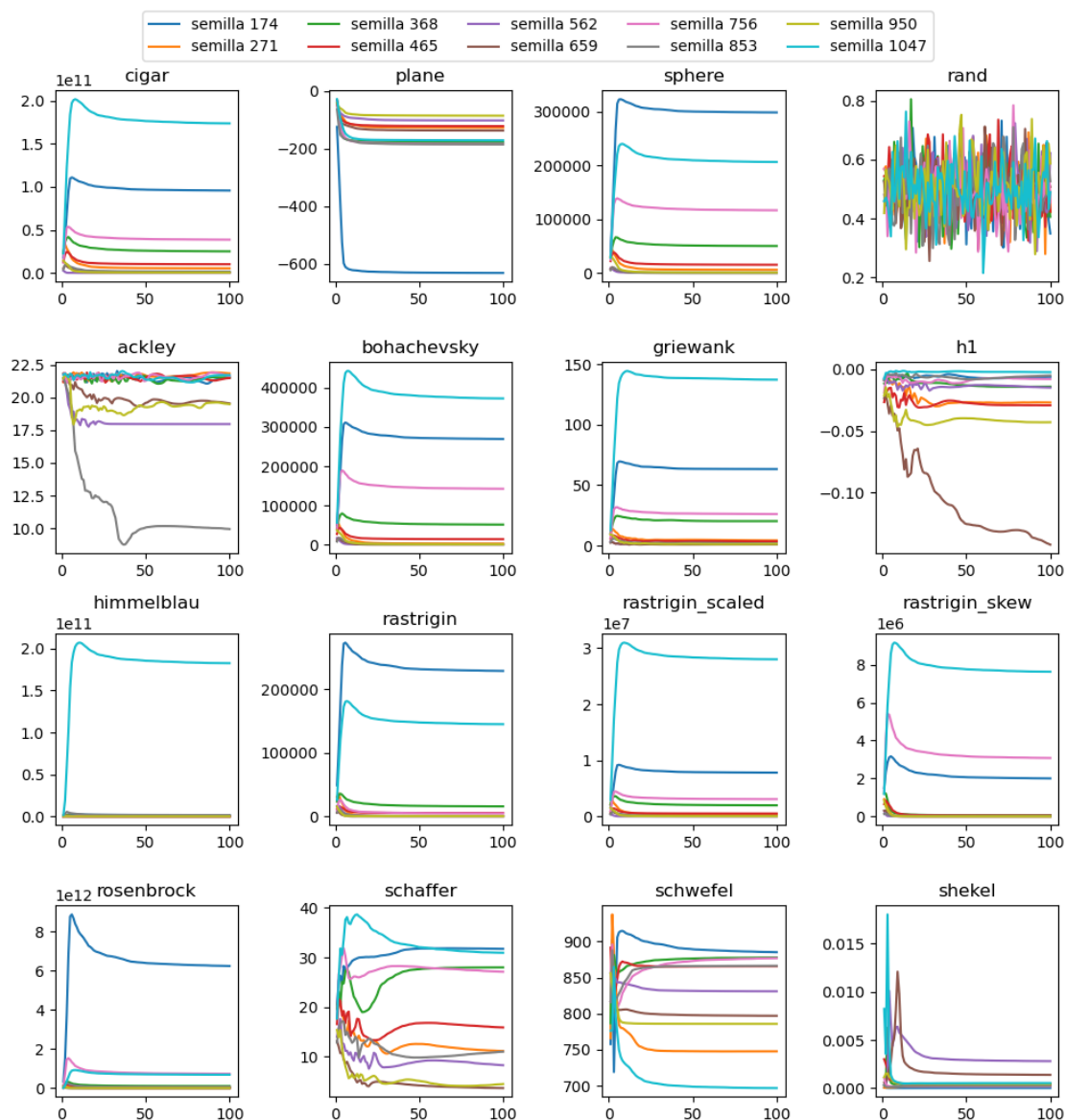
3.1.5 Estabilidad estadística

Los tres algoritmos, fijada una función objetivo, muestran el mismo comportamiento asintótico para varios valores de la semilla. Estos comportamientos no son iguales entre los algoritmos. Se probaron un total de 10 valores distintos con los que inicializar los algoritmos y en líneas generales se vio que el valor de este parámetro afecta en gran medida la convergencia del algoritmo a un valor u otro.

3.1.5.1 GSA

Se puede ver cómo hay valores de las semillas que funcionan mejor para un tipo de funciones frente a otros. Aunque la curva descrita por la media es muy similar para los distintos valores de la semilla su magnitud cambia significativamente. Es el caso de las

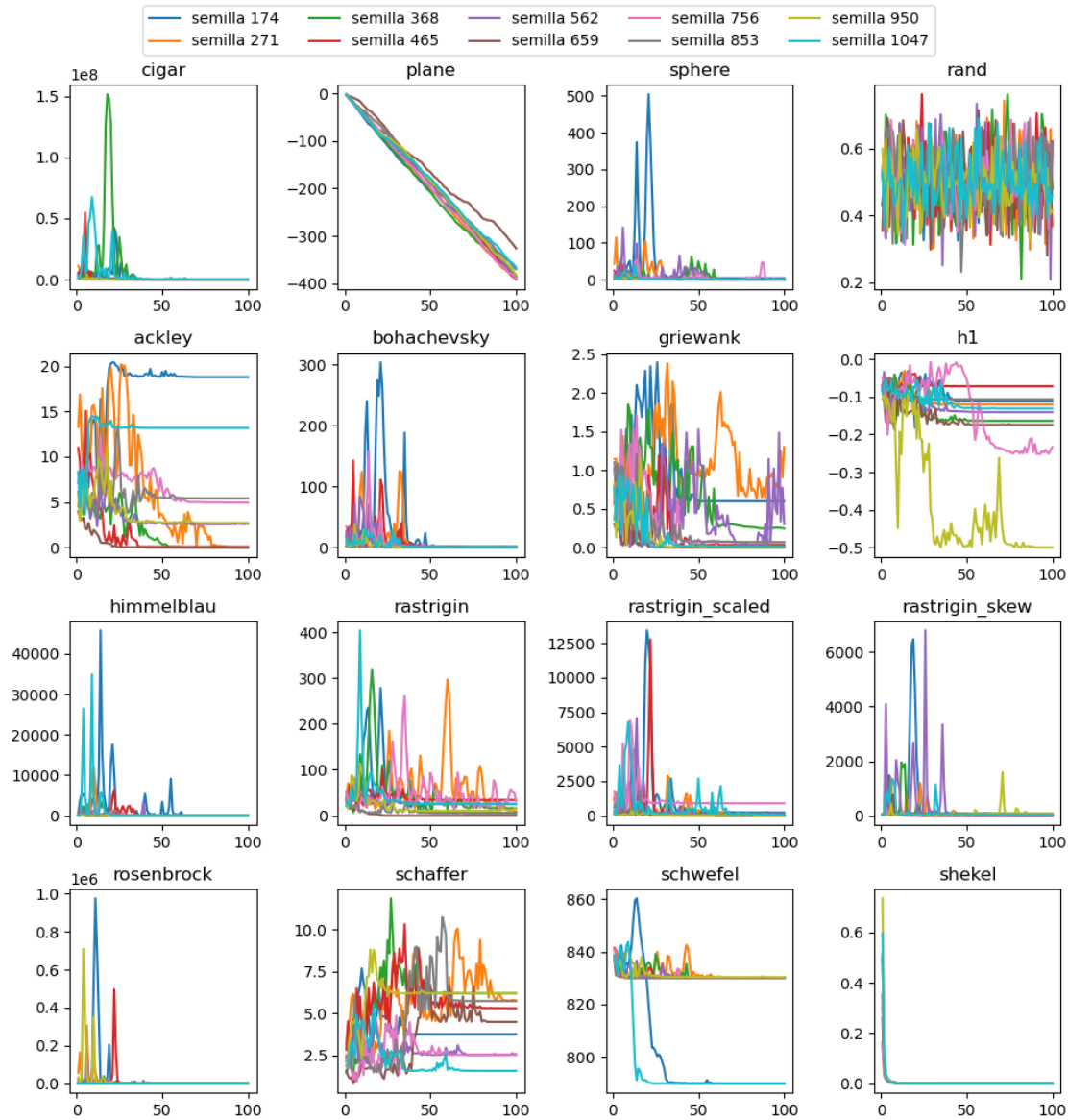
funciones cigar, sphere o bohachevsky, donde el valor de la semilla puede suponer un valor de la función fitness extremadamente alto.



Gráfica 11: Efecto del valor de la semilla en el algoritmo GSA

3.1.5.2 PSO

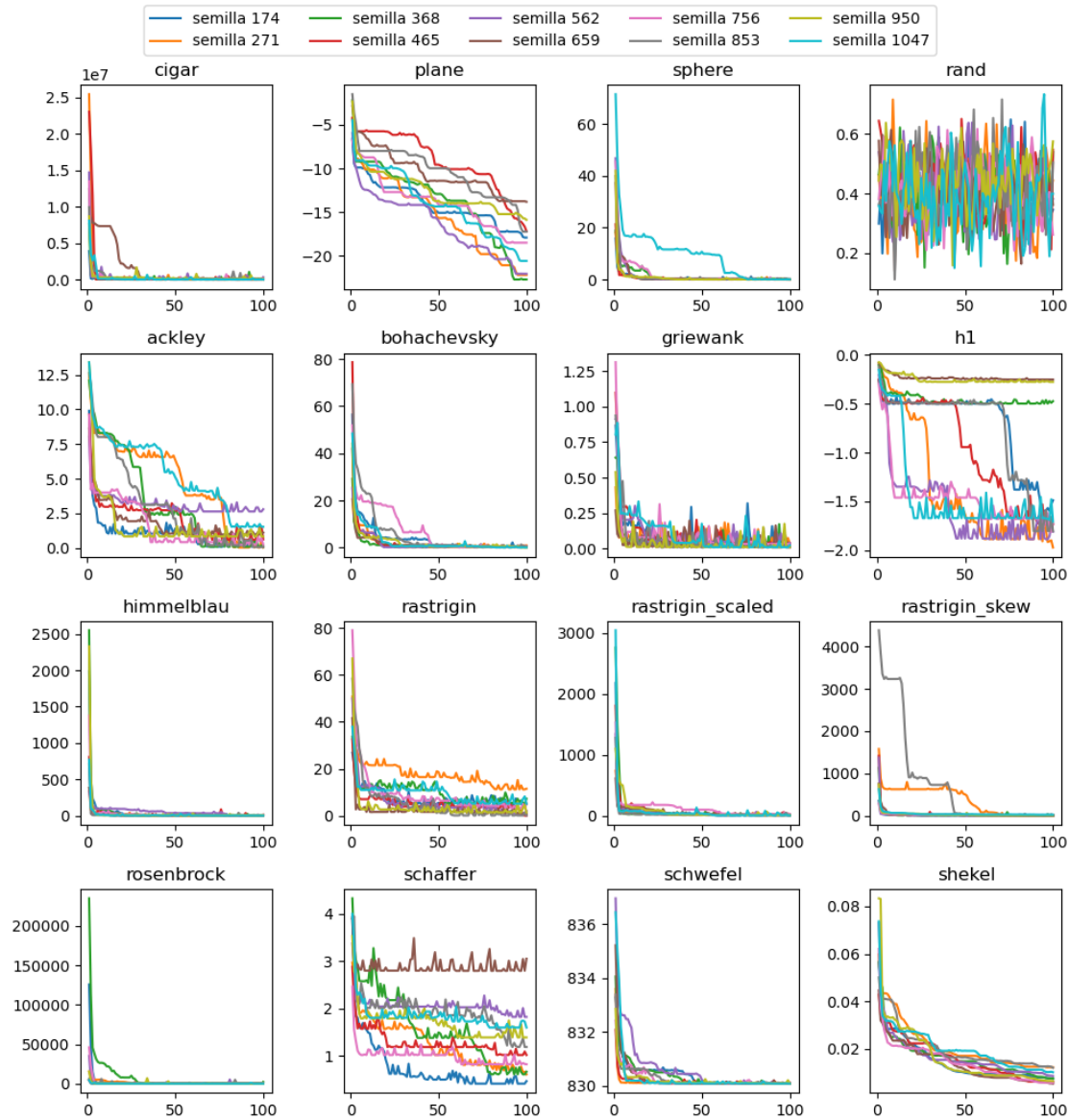
Nuevamente, el valor de la semilla tiene un efecto significativo en el valor al que converge el algoritmo como se ve en la función ackley o schaffer. Aunque en líneas generales, es notable que el valor de la semilla tiene más efecto en la volatilidad del algoritmo en su ejecución que en el valor al que converge.



Gráfica 12: Efecto del valor de la semilla en el algoritmo PSO

3.1.5.3 GA

En el caso de los genéticos no hay tanto impacto del valor de la semilla, todas las ejecuciones parecen converger a un mismo punto. Si bien, según el valor elegido para la semilla la convergencia a este valor se dará más pronto o más tarde en la ejecución.



Gráfica 13: Efecto del valor de la semilla en el algoritmo GA

3.2 Comparación

Se comparan los resultados de los tres algoritmos en función de dos estadísticos, la media y la desviación típica, para un numero de 100 iteraciones fijo para todas las funciones fitness y algoritmos.

Para la mayoría de las funciones el desempeño de PSO es mejor en términos medios. Es capaz de hallar mínimos con más rapidez que los otros dos algoritmos. La segunda mejor opción parece ser GA, en términos de media. Para algunas funciones, concretamente rand, schwefel y shekel se alcanzan valores parecidos.

En cuanto a la desviación estándar depende más de la función objetivo a evaluar que del algoritmo aplicado. Si bien, se pueden notar que PSO posee desviaciones más pequeñas frente a los otros dos algoritmos.

Fitness	Estadístico	GSA	PSO	GA
cigar	media	13005917164.5837	0.0978	5878.1373
	desviación estándar	11810938748.7596	0.0000	0.0000
	generación	100	100	100
plane	media	-139.5131	-395.3225	-18.7130
	desviación estándar	50.5797	0.7653	0.4450
	generación	100	100	100
sphere	media	30191.1452	0.0031	0.0019
	desviación estándar	18140.2691	0.0038	0.0000
	generación	100	100	100
rand	media	0.5266	0.5356	0.4337
	desviación estándar	0.3331	0.2631	0.3147
	generación	100	100	100
ackley	media	21.0860	4.9275	1.4722
	desviación estándar	0.9319	0.0052	0.1891
	generación	100	100	100
bohachevsky	media	30282.3321	0.1072	0.4994
	desviación estándar	24223.4966	0.0000	0.0000
	generación	100	100	100
griewank	media	7.7646	0.5512	0.0087
	desviación estándar	4.5727	0.0027	0.0000
	generación	100	100	100
h1	media	-0.1187	-0.1639	-0.3370
	desviación estándar	0.1483	0.0000	0.0085
	generación	100	100	100
himmelblau	media	6139303.6154	2.1593	1.2641
	desviación estándar	12867265.3035	0.0002	3.0958
	generación	100	100	100
rastrigin	media	3523.1433	39.8441	0.8308
	desviación estándar	4360.0581	0.7353	0.0000
	generación	100	100	100
rastrigin_scaled	media	825694.4591	0.0017	43.4703
	desviación estándar	799037.8032	0.0003	65.9240

Fitness	Estadístico	GSA	PSO	GA
rastrigin_skew	generación	100	100	100
	media	147886.7924	4.9792	3.9418
	desviación estándar	230484.8807	0.0003	1.8434
rosenbrock	generación	100	100	100
	media	1981183356.1607	27.4510	3.2679
	desviación estándar	3909353943.9123	0.0020	0.0000
schaffer	generación	100	100	100
	media	17.8176	4.1051	0.3635
	desviación estándar	8.5978	0.0001	0.0625
schwefel	generación	100	100	100
	media	784.8356	830.0755	830.0752
	desviación estándar	137.1556	0.0000	0.0000
shekel	generación	100	100	100
	media	0.0019	0.0000	0.0054
	desviación estándar	0.0023	0.0000	0.0002
	generación	100	100	100

Tabla 4: Comparación de la media y desviación de los tres algoritmos para las funciones objetivo

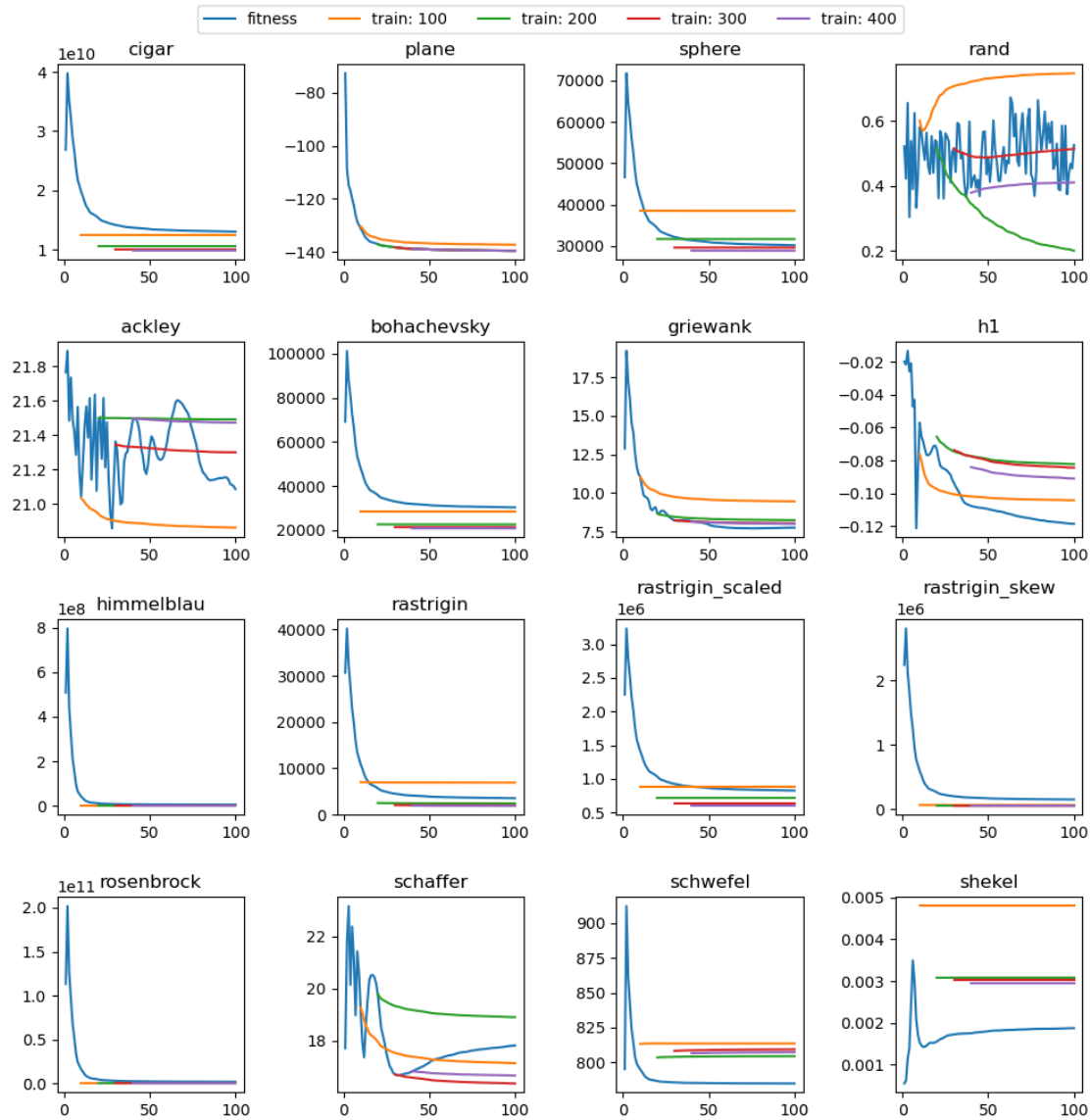
4 Modelos Surrogados

Los modelos surrogados suelen utilizarse en problemas en los que la función objetivo es difícil o costosa de evaluar. Este tipo de modelos actúan como sustituto de la función objetivo siendo más fáciles de evaluar a costa de cometer cierto error sobre el valor real de la función objetivo. Estos modelos requieren ciertas evaluaciones de la función objetivo para obtener suficientes datos con los que entrenar un modelo de aprendizaje automático. El tamaño y la calidad de esta muestra tiene un impacto directo en como de fiables sus predicciones.

4.1 Efecto del tamaño de la muestra

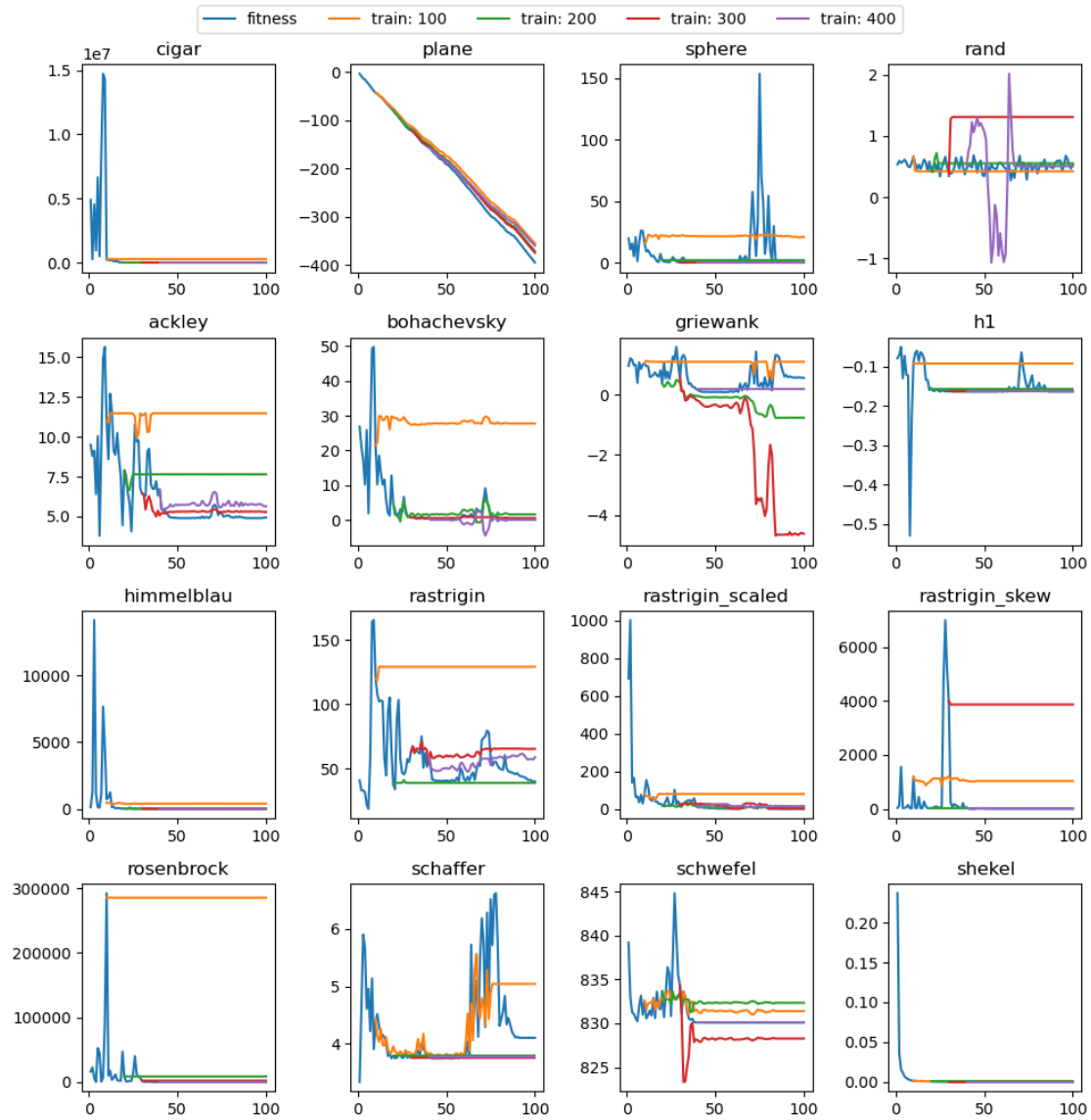
Comenzamos estudiando el efecto del tamaño de la muestra tomada para ajustar el modelo.

Hay funciones que se benefician de un tamaño pequeño de la población, siendo la diferencia entre el valor real y el estimado por el modelo muy pequeño y pudiendo utilizarse desde iteraciones muy tempranas de la ejecución. Para GSA es el caso de las funciones cigar, bohachevsky o h1, donde con una muestra de 100 individuos es suficiente como para aproximarse al valor de la función objetivo.



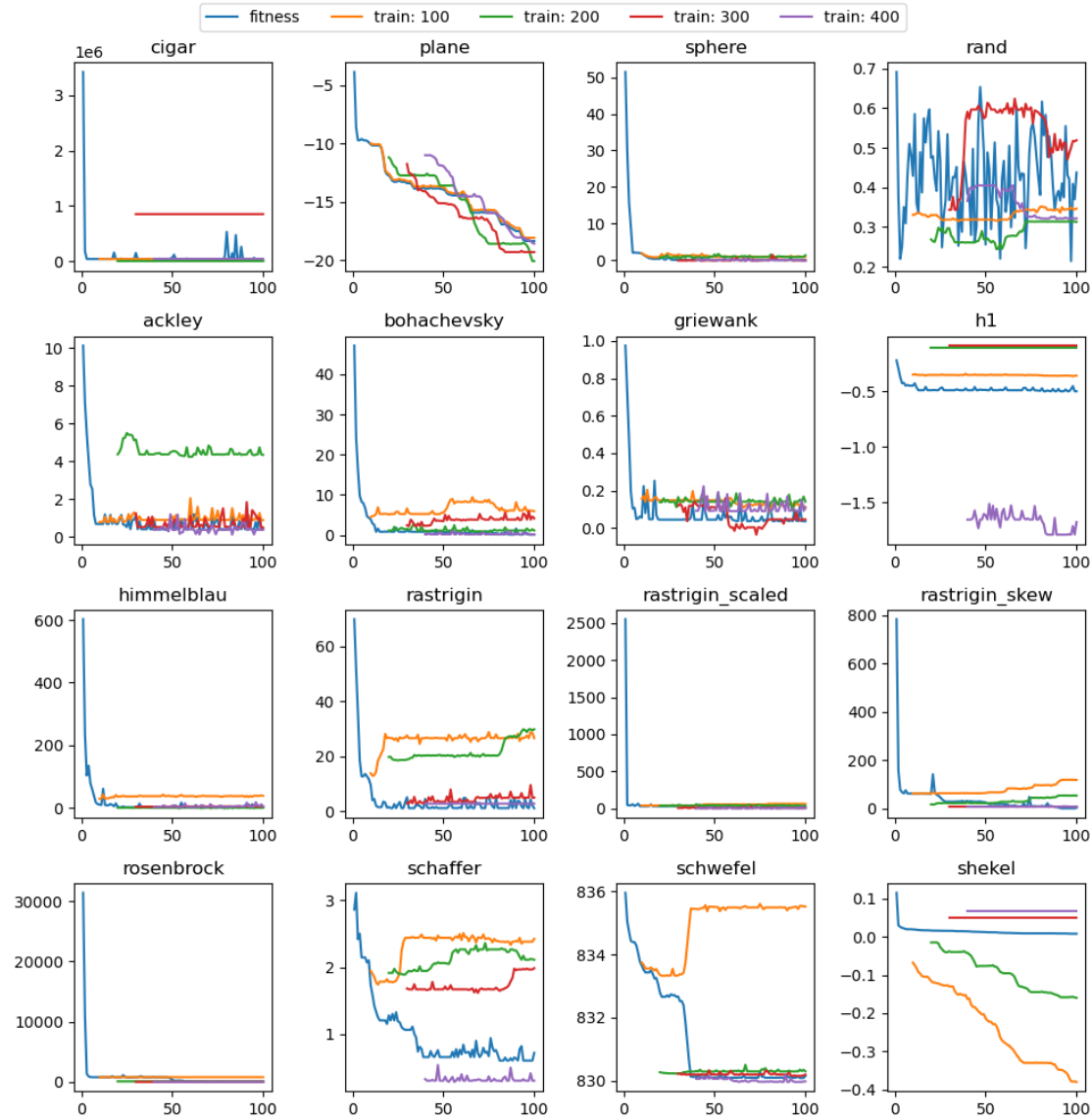
Gráfica 14: evolución de la predicción del modelo surrogado en GSA según el tamaño de la muestra

En el caso de PSO, el general, a mayor tamaño de la muestra mejor predicción de la función objetivo se obtiene.



Gráfica 15: evolución de la predicción del modelo surrogado en PSO según el tamaño de la muestra

El algoritmo genético



Gráfica 16: evolución de la predicción del modelo surrogado en GA según el tamaño de la muestra

4.2 Error del surrogado

Se compara el error relativo del surrogado en función del tamaño del conjunto de entrenamiento para las distintas funciones objetivo. En esta tabla se muestran los errores que se cometen con el algoritmo GSA, el análisis es equivalente para los otros dos algoritmos. En líneas generales podemos ver una disminución del error conforme aumenta el tamaño del conjunto.

fitness	100	200	300	400
cigar	1.912245	1.595961	1.498249	1.459856
plane	0.029427	0.000582	0.000600	0.000607

fitness	100	200	300	400
sphere	1.913869	1.528646	1.412798	1.363853
rand	0.030075	0.034881	0.028822	0.025703
ackley	0.024003	0.018050	0.021631	0.016086
bohachevsky	1.603962	1.229918	1.146236	1.106901
griewank	0.083929	0.039142	0.029983	0.028764
h1	0.012008	0.011147	0.009922	0.009721
himmelblau	0.438244	0.379217	0.376617	0.371434
rastrigin	4.137958	1.049664	0.739565	0.614589
rastrigin_scaled	1.503079	1.158682	1.010444	0.944839
rastrigin_skew	1.799102	1.459580	1.367470	1.327387
rosenbrock	4731.020086	570.978370	267.246359	192.699108
schaffer	0.253592	0.324429	0.163628	0.069678
schwefel	0.120739	0.085511	0.062629	0.043022
shekel	0.000347	0.000244	0.000242	0.000239

Tabla 5: Error relativo del surrogado en GSA según la función objetivo y el tamaño de la población