

# Algoritmos genéticos: la evolución como modelo matemático

**Raúl Corzo Suárez**  
**Francisco Javier Paredes Olay**

**D**URANTE MILLONES de años los seres vivos se han encontrado con numerosas situaciones adversas, es decir, con una enorme cantidad de problemas que han tenido que ir solucionando poco a poco mediante sucesivas adaptaciones. El éxito de la vida en innumerables entornos no es sino el reflejo de que los seres vivos han encontrado soluciones para los distintos problemas con los que se han enfrentado. Son varias las cuestiones que podemos plantearnos en relación a esta cuestión: ¿cuál es el mecanismo que ha permitido la supervivencia de los seres vivos en ambientes tan distintos?, ¿existe algún algoritmo matemático que subyazca en el mismo?, en este caso, ¿podría ser aplicable a otras situaciones y problemas? Los algoritmos genéticos son una de las herramientas que han nacido para responder a estas cuestiones.

El advenimiento de los computadores ha supuesto la posibilidad de enfrentarse a problemas que hace tiempo parecían inabordables. Se precisan, no obstante, modelos adecuados que sean capaces de utilizar acertadamente la fuerza de cálculo tan potente de los ordenadores. Dentro de este campo se han desarrollado métodos como el de Montecarlo, el de cristalización simulada, las redes neuronales o los propios algoritmos genéticos. En el presente trabajo nos centramos en este último modelo en base al éxito que está obteniendo, sobre todo, en el campo matemático de la optimización, tanto a nivel industrial como a nivel científico.

## Un poco de historia

En 1831, a los 22 años de edad, un joven naturalista se embarca en el velero *Beagle* que le llevará a realizar un apasionante viaje de más de tres años por las costas de Sudamérica. Con apenas una libreta por compañía, el joven observa detenidamente la fauna y la flora de las selvas tropicales, de los fríos desiertos de la Tierra de Fuego, de las islas... y una idea va formándose en su cabeza. Acabado el viaje llega el regreso al hogar y, algunos años más tarde, el naturalista publica un libro cuya primera edición se agotó el mismo día de su publicación y que llevaba por título *Sobre el origen de las especies por medio de la selección natural* y cuyo autor era un tal Charles Darwin...

Los postulados básicos en los que descansa la teoría evolutiva de Darwin son los siguientes:

1. El número de individuos de la mayoría de las poblaciones naturales permanece constante debido a la limitación de los recursos y a la competencia por su uso.
2. Dentro de una población los individuos difieren entre sí y estas diferencias pueden ser positivas o negativas para poder competir por los recursos.
3. El resultado de esta competencia es la selección natural: sobrevivirán los individuos mejor adaptados.

Mucho ha llovido desde que aquel joven iniciara aquel viaje hacia la gloria y, sobre todo, mucho se ha desarrollado la Biología Molecular desde entonces. La teoría neodarwinista, basada en los estudios de Darwin y fundamentada en los conocimientos actuales sobre Genética y Biología Molecular se basa en los siguientes principios:

1. Para que una población evolucione deben existir diferencias entre los individuos de la misma y los responsables de estas diferencias son los genes.
2. Algunos de los individuos estarán mejor adaptados al medio que el resto.
3. Los individuos mejor adaptados contribuyen con más descendientes a las siguientes generaciones.

## Un poco de Genética

Nuestro punto de partida es una población que estará formada por individuos que, lógicamente, presentarán alguna diferencia entre ellos. Vamos a suponer que estas diferencias entre individuos se deben a diferencias entre sus genes, es decir, diferencias en el *genotipo*, que es el conjunto de genes de un individuo. Definiremos también el *fenotipo* como la expresión observable del genotipo.

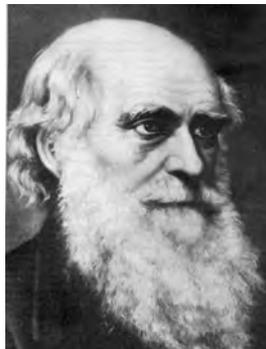
Existen diversos mecanismos cuya acción supone alteraciones en el genotipo, entre los que destacaremos dos: mutación y cruce.

### a) *Mutación*:

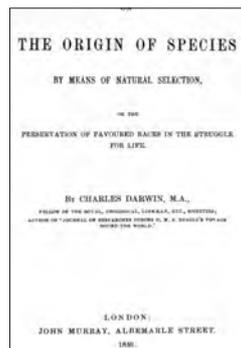
La mutación se define como la alteración de uno o más genes. Esto hace cambiar las características del individuo y, por lo tanto, su nivel de adaptación. Por lo general las mutaciones suelen empeorar la adaptación del individuo al medio, pero algunas de ellas la mejoran. Esta es una operación muy importante, puesto que a pesar de su baja probabilidad de mejorar al individuo, permite la introducción de nuevas características en la población.

### b) *Cruce*:

La operación de cruce consiste en la combinación de dos cromosomas para crear uno o más individuos nuevos. En este caso, los individuos mejor adaptados serán los que más probabilidades tengan de cruzarse con



Charles Darwin



Portada de la primera edición de *El origen de las especies*

otros. El tipo de cruce más utilizado es el cruce puntual, que consiste en «partir» por un mismo punto los dos cromosomas a cruzar e intercambiar sus partes obteniendo así dos nuevos individuos. Otro tipo de cruce, es el cruce uniforme, que recorre los genes de cada uno de los cromosomas y los intercambia al azar.

## Algoritmos genéticos

Los algoritmos genéticos (AG) trabajan a partir de estas equivalencias:

- *Gen*: un conjunto finito de bits.
- *Genotipo* o *cromosoma*: el conjunto de los genes, es decir, todas las cadenas de bits.
- *Mutación puntual*: alteración en un bit de un gen al azar.
- *Cruce*: intercambio de bits entre las cadenas de bits.
- *Función objetivo*: entorno al que deben adaptarse los individuos para sobrevivir.
- *Factor de mutación*: fracción de individuos que sufren mutación.
- *Factor de intercambio*: fracción de individuos que sufren intercambio de bits.

Inicialmente, se parte de una población que presente individuos con distintos genotipos, es decir, que tengan distintas cadenas de bits. Tras la aplicación de las operaciones de mutación y cruce sobre un porcentaje de nuestra población inicial, se obtiene una serie de nuevos individuos que pasan a formar parte de la población. Para lograr que la población se mantenga constante en número de individuos y no crezca en progresión geométrica, se aplica lo que se denominan *políticas de dispersión*.

Dicho de manera un tanto violenta, una política de dispersión decide qué individuos han de morir para evitar el crecimiento de la población. Normalmente no se suelen aplicar *políticas de envejecimiento*, por lo que la supervivencia de un individuo se decide por

su nivel de adaptación sin que influya su edad en ello.

Las dos políticas de dispersión más utilizadas en los AG son:

a) *Dispersión elitista:*

Los nuevos individuos se comparan con los peores individuos de la población, y sobreviven los mejor adaptados a la función objetivo.

b) *Dispersión pseudoelitista:*

La mitad de los nuevos individuos se comparan con los peores de la población, la otra mitad se compara con los mejores y se seleccionan a los mejor adaptados.

Tras estas operaciones se ha concluido lo que se llama una generación. Ahora tenemos una nueva población en la que muy posiblemente tengamos individuos que se adapten mejor a la función objetivo. Tras esto, sólo queda repetir de nuevo las operaciones anteriores hasta conseguir, al menos, un individuo que se adapte perfectamente a las función objetivo, es decir, la solución a nuestro problema.

## El problema de las 8 reinas

La filosofía y desarrollo de un AG se entienden mejor con un ejemplo. En este caso nos hemos decidido por tratar de resolver el problema de las ocho reinas debido a que la aplicación de los algoritmos genéticos en este caso es muy clarificador en cuanto a su funcionamiento.

El problema de las 8 reinas consiste en colocar sobre un tablero de ajedrez ocho reinas sin que ninguna de ellas esté amenazada por ninguna otra. El problema original (que fue investigado por C. F. Gauss en 1850 sin llegar a resolverlo completamente) consiste en hallar todas las posibles soluciones. En nuestro caso, no obstante, nos vamos a limitar a buscar alguna de las soluciones posibles. Como comentario diremos que existen 92 soluciones del pro-

*El problema de las 8 reinas consiste en colocar sobre un tablero de ajedrez ocho reinas sin que ninguna de ellas esté amenazada por ninguna otra. El problema original (que fue investigado por C. F. Gauss en 1850 sin llegar a resolverlo completamente) consiste en hallar todas las posibles soluciones.*

blema de las 8 reinas, aunque sólo son 12 las sustancialmente diferentes, pues el resto son soluciones simétricas de estas (Wirth, 1980). Las combinaciones posibles de reinas sobre el tablero, con la restricción impuesta de colocar cada reina en una columna, son 16.777.216 (es decir,  $8^8$ ).

## Consideraciones generales de los AG para el problema de las 8 reinas

Un tablero de ajedrez consta de 8 columnas y 8 filas. Por otro lado, la reina se puede desplazar tantas casillas como quiera en cualquiera de los ocho sentidos (Figura 1).

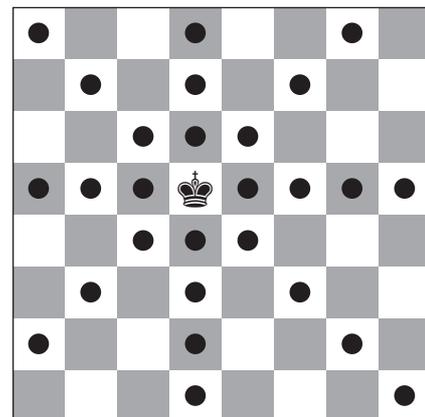


Figura 1. Movimientos de la reina

En el problema que nos ocupa cualquier solución que obtengamos debe tener únicamente una reina por columna. Partiendo de esta consideración, representaremos un individuo mediante un vector de ocho elementos en el que cada elemento representa una de las columnas del tablero, y el valor del elemento representa la fila en la que se encuentra la reina dentro de esa columna. Numerando las filas de 0 a 7 (en lugar de 1 a 8) la figura 2 recoge el ejemplo de un individuo cuya representación vectorial sería (2,4,0,6,1,3,5,1).

¿A qué se debe que empecemos numerando las filas con el número 0? Pues a que el AG trabaja con estos números pero en formato binario, por lo que para representar ocho valores diferentes podemos utilizar tres *bits* ( $2^3=8$ ). Así, nuestro cromosoma constará de un total de  $8 \times 3 = 24$  *bits*. En el caso del ejemplo anterior, el vector quedaría representado por la siguiente cadena de bits: 010 100 000 110 001 011 101 001. Obsérvese que numerando las filas de 0 a 7 en lugar de 1 a 8 aprovechamos al máximo los 3 bits (ya que el número 8 en binario es 1000, número de 4 *bits*).

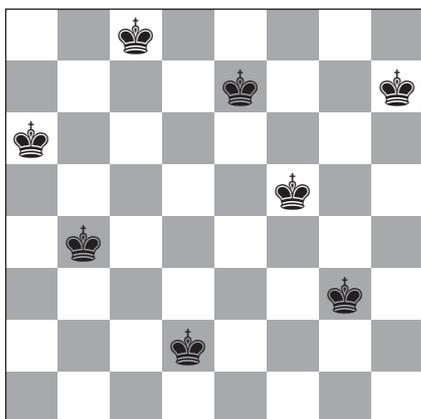


Figura 2. Ejemplo de una disposición al azar de 8 reinas  
Cromosoma: 010 100 000 110 001 011 101 001

De esta manera, ya hemos definido el genotipo de un individuo, que consistirá en una cadena de veinticuatro bits. La interpretación de estos datos, es decir, la disposición de las reinas sobre el tablero, sería el fenotipo.

El siguiente paso es definir la función de ajuste. Esta nos dará una valoración de lo bien que nuestro individuo se adapta al medio, es decir, lo cerca o lejos que se encuentra de la solución al problema. La elección de esta función, junto con la definición del cromosoma, son los dos pasos más críticos, y son los que marcarán en gran medida el éxito del algoritmo así como su rapidez.

Nuestra función de ajuste debe tratar de penalizar aquellas disposiciones de las damas en las cuales una fila o una diagonal estén ocupadas por más de una dama. Sabemos que una reina situada en la fila  $x$  y en la columna  $y$  ocupará también las diagonales  $x + y$  y  $x - y$ . En el tablero de ajedrez existen ocho filas, ocho columnas, quince diagonales en una dirección y quince en la otra.

Nuestra función de ajuste utiliza tres vectores. Uno con ocho elementos representando las filas y los otros dos, con quince elementos cada uno, que representan las diagonales. Inicialmente todos los elementos valen  $-1$ . Para evaluar una posible solución atendemos a las posiciones de las damas. A partir de las coordenadas de cada una de ellas obtenemos tres valores: el número de fila que ocupa, y los números de las dos diagonales. Con estos valores incrementamos en una unidad los elementos que le correspondan en los vectores anteriormente mencionados. Tras estas operaciones, realizamos la suma de todos los elementos mayores que cero, siendo la mejor solución aquella que se encuentre más cercana al valor cero, pues sólo la coincidencia de dos o más piezas en una fila o diagonal producirá valores mayores que cero en el vector correspondiente. La función de ajuste para el ejemplo

*Nuestra función de ajuste debe tratar de penalizar aquellas disposiciones de las damas en las cuales una fila o una diagonal estén ocupadas por más de una dama.*

anterior sería 6, que se corresponde con la coincidencia de damas en una fila, en dos diagonales en un sentido y en tres diagonales en el otro, tal y como puede observarse en la figura 3.

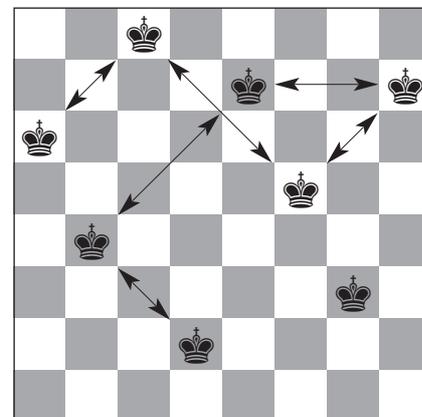


Figura 3. Representación de reinas coincidentes en filas y diagonales

### Condiciones de trabajo del AG

Para la resolución del problema se realizó un programa en lenguaje *C* sobre un *Pentium 233Mhz* con 64Mb RAM bajo sistema operativo *Linux*. Las librerías de AG utilizadas fueron *GALIB v2.4.4* by Matthew Wall at the Massachusetts Institute of Technology. Tras varias pruebas iniciales se optó por elegir las siguientes características de los AG: en cuanto a los operadores que actúan sobre los individuos de la población, en nuestro ejemplo hemos utilizado la mutación simple y el cruce puntual. El tipo de dispersión elegida fue la elitista. Inicialmente se partió de una población de 20 individuos. El factor de la mutación fue de 0,5 y el de recombinación de 0,5. Se programó una supervivencia de 10 individuos por generación.

Con estos datos, podemos seguir cómo ha sido la ejecución del programa mediante las figuras 4, 5 y 6.

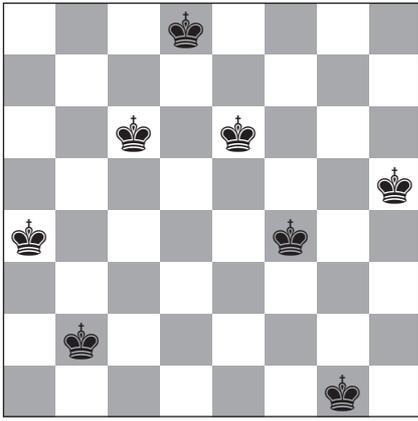


Figura 4. Individuo mejor adaptado de la generación: 1  
 Cromosoma: 100 110 010 000 010 100 111 011  
 Función de ajuste: 3

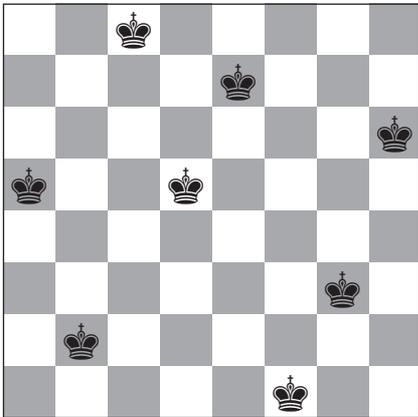


Figura 5: Individuo mejor adaptado de la generación: 22  
 Cromosoma: 011 110 000 011 001 111 101 010  
 Función de ajuste: 1

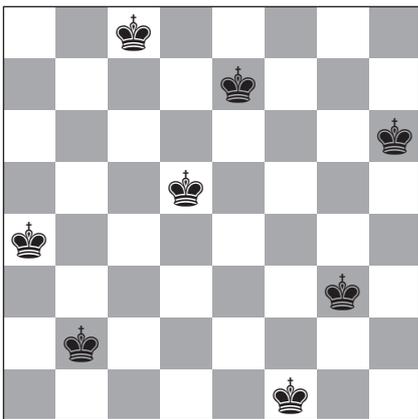


Figura 6: Individuo mejor adaptado de la generación: 23  
 Cromosoma: 100 110 000 011 001 111 101 010  
 Función de ajuste: 0 (solución)

Es decir, en la generación 23 ya ha aparecido algún individuo totalmente adaptado que representa la solución a nuestro problema. Es interesante destacar que si ejecutamos de nuevo el programa podemos obtener soluciones distintas a la mostrada en la figura 6 aun partiendo de los mismos datos.

## Otras aplicaciones de los algoritmos genéticos

### Aplicaciones industriales

Los algoritmos genéticos están teniendo una importante aplicación en diversas industrias debido a que la esencia de determinados problemas es de tal naturaleza que son difícilmente abordables desde la matemática convencional. Un problema típico lo constituye la obtención de varias piezas metálicas de diversos tamaños a partir de una viga metálica optimando al máximo su aprovechamiento, es decir, que el material sobrante sea el menor posible. El planteamiento y resolución de un problema semejante mediante tratamientos matemáticos convencionales es realmente complicado. Sin embargo, los algoritmos genéticos han aportado brillantes soluciones en este campo (Paredes, 1995).

Otro ejemplo de aplicación exitosa de algoritmos genéticos en el campo de la industria lo constituye el control de la distancia de paso del acero fundido entre los cilindros de un molino de rodillos utilizado en la industria metalúrgica para dar un espesor concreto a las láminas de acero. Es conocido que la base de este proceso descansa en el conocimiento de muchos parámetros que deben ser ajustados empíricamente. Debido a que los modelos matemáticos de tratamiento de los mismos descansan en funciones no lineales, la determinación de los mismos es bastante complicada. Los algoritmos genéticos, sin embargo, han resultado ser una herramienta tremendamente interesante para resolver este problema (Paredes, 1996).

### Aplicaciones científicas: difracción de rayos-x

La técnica de difracción de rayos-X de monocristal se ha venido utilizando con el propósito de la determinación de las posiciones atómicas en las estructuras cristalinas. La resolución de una estructura típica (Corzo-Suárez, 1997) supone enfrentarse a un sistema de 3.000 ecuaciones no lineales con 300 incógnitas. Estos sistemas de ecuaciones no lineales son demasiado complicados, por lo que en cristalografía no se afronta la resolución direc-

ta del sistema, sino que vienen utilizando ingeniosas transformaciones de estas ecuaciones y haciendo uso de numerosas relaciones matemáticas y probabilísticas (Sands, 1974). En cualquiera de los casos, la solución conlleva un proceso iterativo relativamente largo y tedioso en el que se hace imprescindible poseer un elevado conocimiento de química cristalográfica y en el que no está garantizado el éxito del estudio, ya que los métodos usados permiten abordar el problema pero a costa de perder parte de la información experimental en lo que se ha venido a denominar el problema de las fases, que constituye el principal problema de la cristalografía estructural. La aplicación de los algoritmos genéticos en este campo, aún en pañales hoy día, está suponiendo un importante avance al permitir trabajar con el sistema de ecuaciones original y manejar de esta forma toda la información.

**Raúl Corzo**  
Departamento  
de Química-Física y Analítica.  
Universidad de Oviedo  
**Francisco Javier Paredes**  
EDS España.  
Oviedo

## Bibliografía

- CORZO-SUÁREZ, R. y otros, (1997): «Redetermination of Bis(2-aminopyrimidinium) Tetrabromocopper(II) at 200 and 100 K», *Acta Cryst.*, C53, 1786-1789.
- PAREDES, J. y otros (1995): «Beam cutting optimization system through genetic algorithms», *World Congress on Neural Networks (17-7-1995)*, Washington.
- PAREDES, J. y otros (1996): «Width control in semicontinuous rolling mills through evolutionary strategies», *World Congress on Neural Networks (17-7-1995)*, Washington.
- SANDS, D. E. (1974): *Introducción a la Cristalografía*, Reverté, Barcelona.
- WIRTH, N. (1980): *Algoritmos + Estructuras de Datos = Programas*, Ediciones del Castillo, Madrid.



*Espiral en el paraíso matemático*  
XI Olimpiada Matemática Nacional  
Equipo 1



*Semáforo geométrico*  
XI Olimpiada Matemática Nacional  
Equipo 2  
(En blanco y negro pierde mucho,  
hay que imaginar los colores)



*Geometría por un tubo*  
XI Olimpiada Matemática Nacional  
Equipo 3