

Estrategias Evolutivas

Oscar Hernández Constantino
constantino92@ciencias.unam.mx



Variantes del Algoritmo Evolutivo

Variantes Históricas:

- Algoritmo Genético
- **Estrategias Evolutivas**
- Programación Evolutiva
- Programación Genética

Versiones más recientes:

- Evolución Diferencial
SHADE
- EDAs
- Algoritmos Bio-Inspirados
PSO - ACO
- Híbridos
 - Memético
 - Rutas Encadenas (Path Relinking)

Estrategias Evolutivas

Inicios

- En la década de 1960, tres estudiantes de la Universidad Tecnológica de Berlín intentaban encontrar formas óptimas para minimizar la resistencia del aire.
 - **Ingo Rechenberg**
 - Hans-Paul Schwefel
 - Peter Bienert
- Tuvieron dificultades para resolver de manera analítica, por lo cual probaron cambios aleatorios, seleccionando los que mejor funcionaban.

Estrategias Evolutivas



- En 1964, Ingo Rechenberg hizo la primera publicación de estrategias evolutivas.
- Sus implementaciones eran experimentales.
- Después de obtener su doctorado, publicó su trabajo en un libro en 1973.

Estrategia Evolutiva (1+1)

- El algoritmo original de ES (Evolutionary Strategy) **inicializa una solución** candidata (vector aleatorio x) y evalúa su aptitud, $f(x)$
- La solución candidata es **mutada** y se evalúa el individuo resultante
- El mejor de los dos individuos (padre e hijo) se **selecciona** como punto de inicio para la siguiente generación.
- La idea original fue diseñada para problemas discretos, usando mutaciones pequeñas, por tanto tiende a quedarse atrapado en un óptimo local.
Posteriormente, el algoritmo se modificó para usar mutaciones en espacios continuos.

Pseudocódigo de la Estrategia Evolutiva (1+1)

1. Inicializar la varianza σ^2 , no negativa, para la mutación
2. $x_0 \leftarrow$ individuo generado de manera aleatorio
3. Mientras NO se cumplan condiciones de término
 - a. Generar un vector aleatorio a $r_i \sim N(0, \sigma^2)$ para $i \in [1, n]$
 - b. $x_1 \leftarrow x_0 + r$
 - c. Si x_1 es mejor que x_0 entonces

$$x_0 \leftarrow x_1$$

siguiente generación

Estrategia Evolutiva (1+1)

- Cada generación consiste de 1 padre y 1 hijo, y el mejor individuo sobrevive para la siguiente generación
- Es similar a las estrategias de búsqueda local (hill climbing, escalando la colina)
- El siguiente teorema garantiza que (1+1)-ES eventualmente encuentra el máximo de $f(x)$:

Teorema: Si $f(x)$ es una función continua definida en un dominio cerrado con un óptimo global $f^*(x)$, entonces:

$$\lim_{t \rightarrow \infty} f(x) = f^*(x)$$

t es el número de generación

Estrategia Evolutiva (1+1)

- La varianza σ^2 es un parámetro que se debe ajustar:
 - σ debe ser suficientemente grande para que las mutaciones puedan alcanzar todas las áreas del espacio de búsqueda en un período de tiempo razonable;
 - σ debe ser suficientemente pequeña para que pueda encontrar la solución óptima, dentro de lo deseado.

ES-(1+1), varianza

- Una alternativa puede ser disminuir σ durante el proceso de búsqueda
- En el pseudocódigo anterior, la mutación es llamada “isotrópica” porque cada elemento de x tiene la misma varianza.

- En la práctica, es posible utilizar mutaciones no isotrópicas:

$$x_1 \leftarrow x_0 + N(0, \Sigma)$$

Σ es una matriz diagonal de $n \times n$ con elementos σ_i , $i \in [1, n]$

- Es posible asignar cada σ_i de manera independiente del dominio del i -ésimo elemento de x y la forma de la función objetivo en cada dimensión

ES-(1+1)

- Rechenberg analizó la ES-(1+1) para problemas de optimización simples y concluyó que el 20% de las mutaciones deberían resultar en mejoras en la función de aptitud $f(x)$
- Si la tasa de mutación es mayor del 20%, las mutaciones son muy pequeñas, lo cual guía hacia pequeñas mejoras, que resultan en convergencia muy lenta
- Si la tasa de mutación es menor del 20%, las mutaciones son muy grandes y eso guía a grandes, pero infrecuentes, mejoras, y también provoca convergencia lenta.

ES-(1+1), Regla de 1/5

Rechenberg propuso la siguiente regla:

En las ES-(1+1), si la tasa de éxito del total de las mutaciones es menor que $\frac{1}{5}$, la desviación estándar σ debería disminuirse. Si la tasa es mayor que $\frac{1}{5}$, la desviación estándar debería incrementarse.

Esto nos deja una pregunta: ¿Por cuánto debería disminuirse o incrementarse?

disminuir la desviación estándar: $\sigma \leftarrow c \sigma$

aumentar la desviación estándar: $\sigma \leftarrow \sigma / c$

Schwefel derivó teóricamente el factor de cambio: $c = 0.817$

Pseudocódigo - ES (1+1) adaptativa

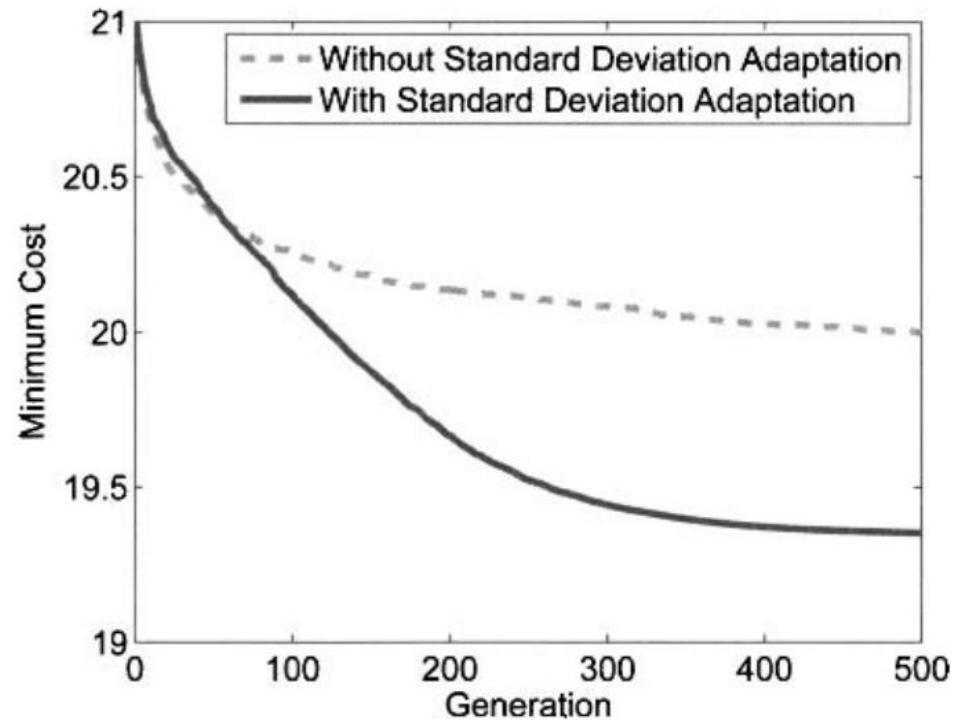
1. Inicializar la varianza σ^2 , no negativa, para la mutación
2. $x_0 \leftarrow$ individuo generado de manera aleatorio
3. Mientras NO se cumplan condiciones de término
 - a. Generar un vector aleatorio a $r_i \sim N(0, \sigma^2)$ para $i \in [1, n]$
 - b. $X_1 \leftarrow X_0 + r$
 - c. Si X_1 es mejor que X_0 entonces $X_0 \leftarrow X_1$
 - d. $\phi \leftarrow$ proporción de mutaciones con éxito durante las últimas G generaciones
 - e. Si $\phi < \frac{1}{5}$
 - i. $\sigma \leftarrow c^2 \sigma$
 - f. Si no
 - i. $\sigma \leftarrow \sigma / c$

Schwefel recomienda:

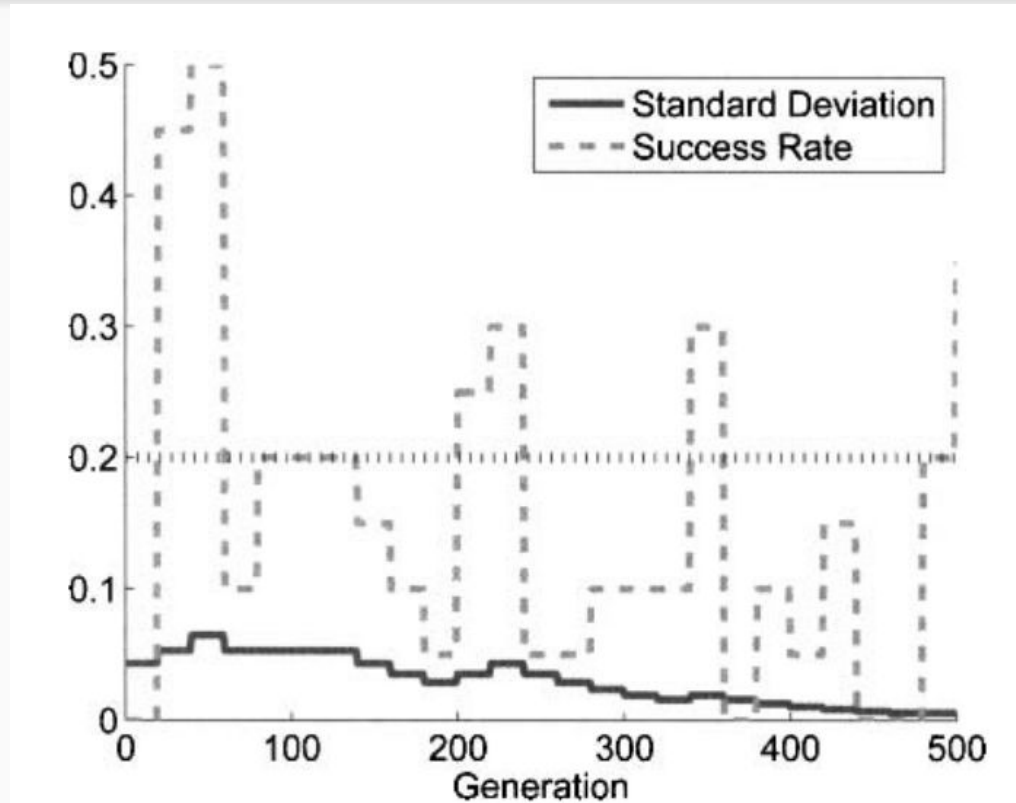
$$G = \min(n, 30)$$

siguiente generación

Ejemplo - Optimización de la función Ackley en dimensión 20



Ejemplo - Optimización de la función Ackley en dimensión 20



Estrategia Evolutiva ($\mu+1$)

- La primera generalización de las ES-(1+1) fueron las ES-($\mu + 1$)
- μ padres son usados en cada generación
- Cada padre tiene asociado un vector σ que controla la magnitud de las mutaciones
- Los padres se combinan para generar un hijo y después el hijo es mutado
- Se eligen a los mejores μ individuos, entre los μ padres y el hijo

[Algoritmo *elitista* de *estado estacionario*]

Pseudocódigo - ES ($\mu+1$)

1. $\{x_k, \sigma_k\} \leftarrow$ inicializar de manera aleatoria, $\sigma_k > 0$, $k \in [1, \mu]$,
2. Mientras NO se cumplan condiciones de término
 - a. Seleccionar a dos padres de la población $\{x_k, \sigma_k\}$
 - b. Usar un método de recombinación para obtener un hijo: $\{x_{\mu+1}, \sigma_{\mu+1}\}$
 - c. Generar un vector aleatorio r desde $N(0, \Sigma = \text{diag}(\sigma_{\mu+1,1}^2, \dots, \sigma_{\mu+1,n}^2))$
 - d. $x_{\mu+1} \leftarrow x_{\mu+1} + r$
 - e. Remover al peor individuo de la población:
 $\{x_k, \sigma_k\} \leftarrow$ los mejores μ individuos de $\{(x_1, \sigma_1), \dots, (x_{\mu+1}, \sigma_{\mu+1})\}$

siguiente generación

Ejemplo - Operadores de Recombinación

Operador de Cruza Sexual Discreta (Cruza Uniforme)

Parent 1:

Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	σ_{11}	σ_{12}	σ_{13}	σ_{14}	σ_{15}
----------	----------	----------	----------	----------	---------------	---------------	---------------	---------------	---------------

Parent 2:

Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}	σ_{21}	σ_{22}	σ_{23}	σ_{24}	σ_{25}
----------	----------	----------	----------	----------	---------------	---------------	---------------	---------------	---------------

Offspring:

Y_{11}	Y_{22}	Y_{13}	Y_{14}	Y_{25}	σ_{21}	σ_{22}	σ_{13}	σ_{14}	σ_{25}
----------	----------	----------	----------	----------	---------------	---------------	---------------	---------------	---------------

Ejemplo - Operadores de Recombinación

Operador de Cruza Sexual Intermedia

Parent 1:

y_{11}	y_{12}	y_{13}
----------	----------	----------

σ_{11}	σ_{12}	σ_{13}
---------------	---------------	---------------

Parent 2:

y_{21}	y_{22}	y_{23}
----------	----------	----------

σ_{21}	σ_{22}	σ_{23}
---------------	---------------	---------------

Offspring:

$\frac{y_{11} + y_{21}}{2}$	$\frac{y_{12} + y_{22}}{2}$	$\frac{y_{13} + y_{23}}{2}$
-----------------------------	-----------------------------	-----------------------------

$\frac{\sigma_{11} + \sigma_{21}}{2}$	$\frac{\sigma_{12} + \sigma_{22}}{2}$	$\frac{\sigma_{13} + \sigma_{23}}{2}$
---------------------------------------	---------------------------------------	---------------------------------------

Ejemplo - Operadores de Recombinación

Operador de Cruza global discreta (Cruza multi-padres)

Individual 1:	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	σ_{11}	σ_{12}	σ_{13}	σ_{14}	σ_{15}
Individual 2:	Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}	σ_{21}	σ_{22}	σ_{23}	σ_{24}	σ_{25}
Individual 3:	Y_{31}	Y_{32}	Y_{33}	Y_{34}	Y_{35}	σ_{31}	σ_{32}	σ_{33}	σ_{34}	σ_{35}
Individual 4:	Y_{41}	Y_{42}	Y_{43}	Y_{44}	Y_{45}	σ_{41}	σ_{42}	σ_{43}	σ_{44}	σ_{45}
Individual 5:	Y_{51}	Y_{52}	Y_{53}	Y_{54}	Y_{55}	σ_{51}	σ_{52}	σ_{53}	σ_{54}	σ_{55}
Offspring:	Y_{11}	Y_{32}	Y_{53}	Y_{14}	Y_{25}	σ_{51}	σ_{32}	σ_{13}	σ_{44}	σ_{45}

Ejemplo - Operadores de Recombinación

Operador de Cruza Global Intermedia

Individual 1:	y_{11}	y_{12}	y_{13}	σ_{11}	σ_{12}	σ_{13}
Individual 2:	y_{21}	y_{22}	y_{23}	σ_{21}	σ_{22}	σ_{23}
Individual 3:	y_{31}	y_{32}	y_{33}	σ_{31}	σ_{32}	σ_{33}
Individual 4:	y_{41}	y_{42}	y_{43}	σ_{41}	σ_{42}	σ_{43}
Individual 5:	y_{51}	y_{52}	y_{53}	σ_{51}	σ_{52}	σ_{53}
Offspring:	$\frac{y_{21} + y_{31}}{2}$	$\frac{y_{12} + y_{42}}{2}$	$\frac{y_{23} + y_{53}}{2}$	$\frac{\sigma_{41} + \sigma_{51}}{2}$	$\frac{\sigma_{22} + \sigma_{42}}{2}$	$\frac{\sigma_{13} + \sigma_{33}}{2}$

Estrategia Evolutiva ($\mu+\lambda$)

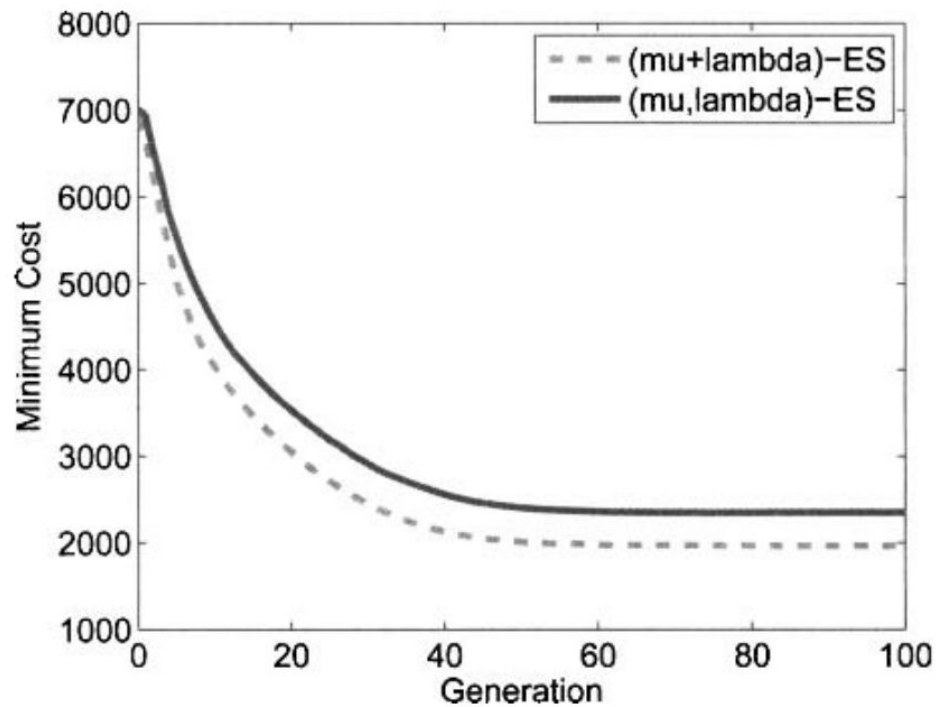
- La siguiente generalización de las estrategias evolutivas fueron las ES- $(\mu + \lambda)$
- Se tiene una población de μ padres
- En cada generación se generan λ hijos
- Se eligen a los mejores μ individuos, entre los μ padres y los λ hijos

[Algoritmo ***elitista generacional***]

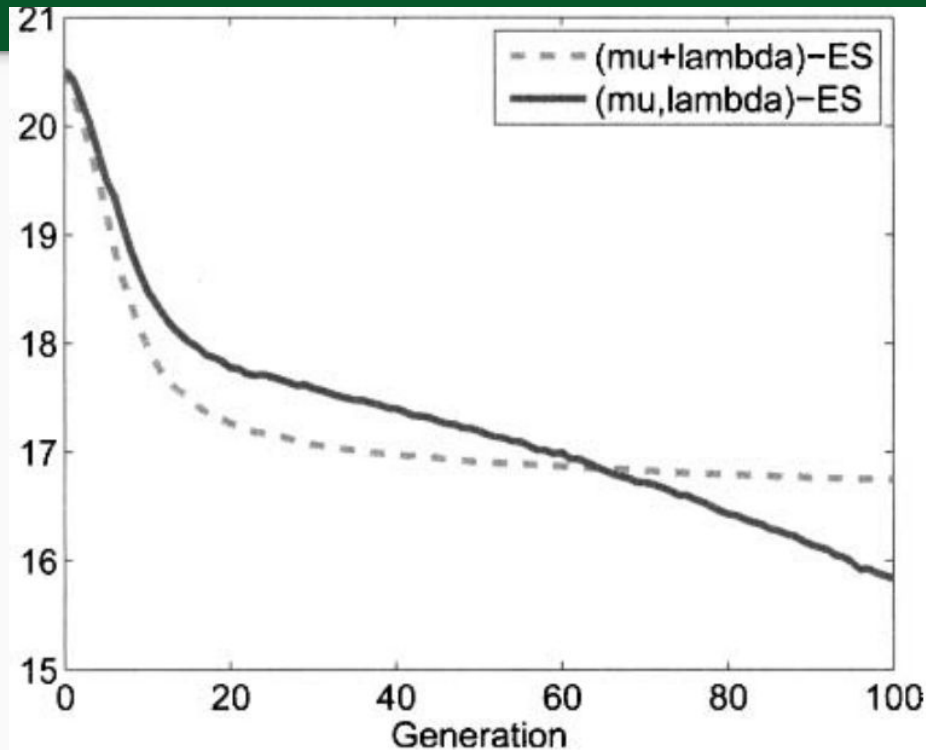
Estrategia Evolutiva (μ, λ)

- Otra generalización de las estrategias evolutivas fueron las ES- (μ, λ)
- Se tiene una población de μ padres
- En cada generación se generan λ hijos
- La población de la siguiente generación se eligen entre los mejores μ individuos de los hijos (ninguno de los padres sobrevive). $\lambda \geq \mu$
- Argumento contra $\lambda > 1$: la explotación de la información se retrasa
- Argumento contra $\mu > 1$: la supervivencia de individuos inferiores retrasa el progreso de las ES

Ejemplo



Schwefel, dimensión 20



Ackley, dimensión 20

Auto Adaptación

- Adaptación en la estrategia (1+1), ajustando de acuerdo a la tasa de éxito
- Se puede generalizar fácilmente a estrategias tipo (1+ λ) ,
 - ¿Cómo se puede aplicar a ($\mu + \lambda$) o (μ, λ)?,
 - No es trivial cuando $\mu > 1$
 - Los hijos están compuestos no solamente de mutaciones, si no también de la combinación de los padres
- Al igual que se mutan las características $\{x_i\}$, es posible mutar los elementos $\{\sigma_i\}$. Por ejemplo:

$$\begin{aligned}\sigma'_i &\leftarrow \sigma_i \exp(\tau' \rho_0 + \tau \rho_i) \\ x'_i &\leftarrow x'_i + \sigma'_i r_i\end{aligned}$$

$\rho_0, \rho_i, r_i \sim N(0,1)$
 τ y τ' son parámetros

Algoritmo Genético vs Estrategias Evolutivas

- Las Estrategias Evolutivas son similares a los Algoritmos Genéticos
- Los AG fueron desarrollados codificando soluciones como cadenas de bits
- Las EE (ES) operan con parámetros continuos
- Los AG hacen énfasis en la recombinación
- En las EE lo más importante es la mutación

CMA - Covariance Matrix Adaptation

- Una de las estrategias evolutivas más exitosas es el algoritmo CMA-ES
- Ajusta la distribución de las mutaciones de acuerdo a la función objetivo
- Intenta ajustar perfectamente para funciones objetivo cuadráticas
- Otras funciones pueden ser aproximadas con una cuadrática que esté cercana al óptimo
- La desventaja es que es una estrategia adaptativa que es complicada para ajustar los parámetros
- Este algoritmo ajusta la matriz de Covarianza para aplicar mutaciones