

RESUMEN

***GENETIC ALGORITHMS &
ENGINEERING DESIGN***

Mitsuo Gen and Runwei Cheng

**RESUMEN
CAPITULOS 1-2**

1. FUNDAMENTOS DE LOS ALGORITMOS GENÉTICOS

1.1 Selección

La selección es la que va a ir dirigiendo la búsqueda, de hay la importancia de la presión de selección. Si la presión de selección es muy alta el algoritmo puede converger prematuramente, mientras que si es muy baja el progreso será mucho más lento de lo estrictamente necesario.

Al principio del algoritmo genético conviene que la presión de selección sea baja para poder explorar un espacio posible de soluciones mucho mayor, y en la fase final de búsqueda conviene que la presión de selección sea alta para aprovechar las regiones con mejores soluciones.

Se distinguen tres temas básicos en la fase de selección:

- Espacio de muestreo.
- Mecanismo de muestreo.
- Probabilidad de selección.

1.1.1 Espacio de muestreo.

El proceso de selección quizás cree una nueva población para la siguiente generación que este basada en todos los padres e hijos o en parte de ellos.

Un espacio de muestreo viene caracterizado por dos elementos fundamentales:

- Tamaño del espacio.
- Ingredientes (padres o hijos).

Sea *pop_size* el tamaño de la población y sea *off_size* el número de hijos generados en esta generación.

El *espacio de muestreo regular* viene caracterizado por:

- Tamaño: *pop_size*.
- Ingredientes: Todos los hijos generados.

El *espacio de muestreo ampliado* viene caracterizado por:

- Tamaño: *pop_size* + *off_size*.
- Ingredientes: Todos los hijos generados y todos los padres.

A) Espacio de muestreo regular.

En el AG de Holland los padres eran sustituidos por los hijos. Es lo que se conoce como sustitución generacional. Como los AG son ciegos por naturaleza., los hijos quizás sean peores que sus padres.

Se han estudiado diversas estrategias para solventar este problema:



- Holland sugirió que cuando se produzca un hijo sustituya a un elemento de la población elegido aleatoriamente y no necesariamente su padre.
- De Jong, propuso la conocida como *crowding strategy*, en la cual cuando un hijo nace un padre es seleccionado para morir. El padre seleccionado debe ser semejante al hijo
- Otra estrategia muy conocida es la del tiro de la ruleta.

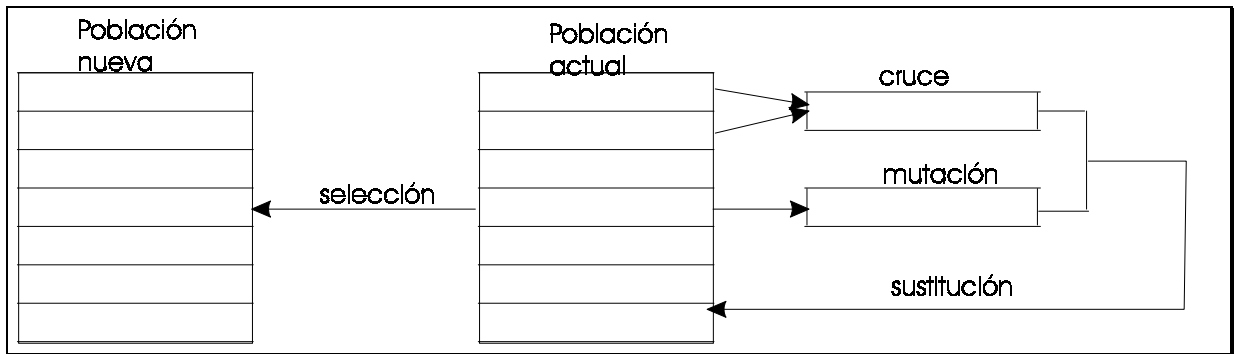


Figura 1: Proceso de selección considerando un espacio de muestreo regular

B) Espacio de muestreo aumentado.

En este caso tanto los padres como los hijos tienen la misma probabilidad de competir para sobrevivir.

Un típico caso es la selección $(\mu+\lambda)$. Con esta estrategia μ padres y λ hijos compiten por sobrevivir. Los μ mejores son sacados de los hijos y de los padres, y se seleccionan como padre para la próxima generación.

Otra estrategia de selección es la conocida como (μ,λ) , que selecciona los mejores μ hijos como padres para la próxima generación, cumpliéndose que $(\mu < \lambda)$.

Estos dos métodos son ambos deterministas.

Una ventaja evidente de esta aproximación es que podemos mejorar el comportamiento del AG aumentando la probabilidad de cruce y la probabilidad de mutación p_c y p_m , sin preocuparnos en que se introduzcan perturbaciones aleatorias, siempre que se trabaje en espacio de muestreo ampliado.

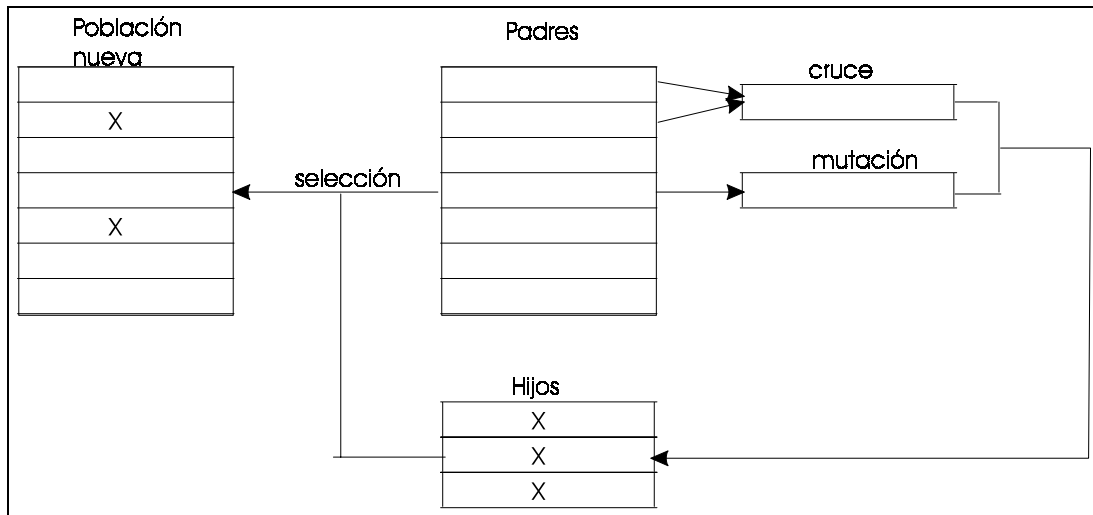


Figura 1: Proceso de selección considerando un espacio de muestreo ampliado.

1.2 Mecanismos de muestreo.

Los mecanismos de muestreo hacen referencia al problema de como seleccionar cromosomas del espacio de muestreo. Existen tres aproximaciones a este problema:

- Muestreo Estocástico.
- Muestreo Determinista.
- Muestreo Mixto

1.2.1 Muestreo Estocástico

La fase de selección se compone de dos partes:

1. Determinar el valor esperado de cada cromosoma.
2. Convertir el valor esperado a número de hijos.

El valor esperado de cada cromosoma se refiere al número real que indica el número medio de hijos que un cromosoma debería recibir. Aquí el proceso de muestreo es utilizado para convertir el valor real esperado a número de hijos.

Un ejemplo típico de este tipo de muestreo es la selección mediante el tiro de una ruleta pop_size veces. Donde la probabilidad de selección es:

$$p_k = \frac{f_k}{\sum_{j=1}^{pop_size} f_j}$$

Baker propuso el conocido como *muestreo universal estocástico* que utiliza un único giro de la ruleta. La ruleta que se utiliza tiene pop_size marcas igualmente espaciadas. El valor esperado e_k para cada cromosoma k se calcula como:

$$e_k = pop_size \cdot p_k$$

```

                                Algoritmo de muestreo estocástico universal
begin
    sum=0;
    ptr=rand( ) ; %esta función devuelve un número aleatorio entre [0,1)
    for k=1 : pop_size
        sum=sum+ek ;
        while(sum>ptr)
            seleccionar cromosoma k ;
            ptr=ptr+1
        end
    end
end
end

```

La base de la consideración de esta aproximación es mantener el número esperado de copias de cada cromosoma en la siguiente generación.

Un aspecto interesante es *la prohibición de mantener duplicados dentro de la población*, existen dos razones para mantener esta estrategia:

1. Prevenir la aparición de supercromosomas mediante el mantenimiento de muchas copias de un mismo individuo, que produce una rápida y prematura convergencia.
2. Mantener la diversidad de la población para que cada generación pueda contener mucha información para la búsqueda genética.

Al descartar elementos duplicados puede suceder que el tamaño salga menor que *pop_size* y sea necesario añadir elementos nuevos provenientes de la población inicial para llegar a *pop_size*.

1.2.2 Muestreo Determinista.

Esta aproximación generalmente selecciona a los mejores *pop_size* cromosomas del espacio de muestreo. La selección $(\mu+\lambda)$ y (μ,λ) pertenecen a esta familia. Estas aproximaciones prohíben la existencia de elementos duplicados

La *selección por truncación* y *selección por bloques* también pertenecen a este método, los cuales clasifican todos los cromosomas de acuerdo a su fitness y seleccionan los mejores como padres.

En selección por truncación se debe definir un umbral T para que los $T\%$ de los mejores cromosomas sean seleccionados y cada uno reciba $100/T$ copias.

En la selección por bloques para una población dada *pop_size*, uno da simplemente s copias a los mejores *pop_size/s* cromosomas.

Ambas implementaciones son idénticas cuando $s=pop_size/T$.



La sustitución generacional puede ser visto como otra versión de la aproximación determinista. Una modificación de este método es sustituir los n peores individuos con n hijos

1.2.3 Probabilidad de selección

En procedimientos de selección proporcionales, la probabilidad de selección de un cromosoma es proporcional a su fitness. Este mecanismo tan simple presenta algunas propiedades no deseadas como:

- En generaciones tempranas existe una tendencia a que unos pocos superindividuos dominen el proceso de selección.
- En las últimas generaciones, la competitividad entre los cromosomas es menor y se tiende a realizar una búsqueda aleatoria.

Los *mecanismos de escalado y de clasificación* surgen con la intención de mitigar estos problemas.

- *Métodos de escalado*: Estos métodos pasa el valor sin procesar de la función de coste (fitness) a algún valor positivo, la probabilidad de supervivencia de cada cromosoma se determina de acuerdo a estos valores.
- *Métodos de clasificación*: ignoran el valor actual de la función de coste y realizan una clasificación de los cromosomas en vez de determinar la probabilidad de supervivencia de los cromosomas.

Para la mayoría de los métodos de escalado, los parámetros de escalado son dependientes del problema.

El fitness escalado f_k' del cromosoma k , será una función del fitness sin procesar f_k .

$$f_k' = g(f_k)$$

Los métodos de escalado se clasifican en dos grupos:

- *Escalado estático*, en este caso la relación entre f_k' y f_k siempre se mantiene constante a lo largo de todo el proceso.
- *Escalado dinámico*, en este caso la relación entre f_k' y f_k varía a lo largo del proceso.

Los métodos de escalado dinámico se subdividen a su vez en dos tipos

1. Los parámetros de escalado son ajustados adaptativamente de acuerdo a la situación del valor de fitness en cada generación con la intención de mantener siempre constante la presión de selección.
2. Los parámetros de escalado son ajustados teniendo como único factor en cuenta el número de generaciones, para ir aumentando la presión de selección de manera adecuada.



Escalado lineal: el escalado lineal ajusta el fitness de todos los cromosomas para que el mejor individuo de la población consiga uno número esperado fijo de hijos y previene una reproducción excesiva de cualquiera de ellos.

$$f_k' = a \cdot f_k + b$$

Los parámetros a y b normalmente son seleccionados para que la media de los cromosomas reciban un hijo en promedio y el mejor cromosoma reciba un número fijado de hijos (normalmente 2). Este método da valores de fitness que pueden ser negativos y con los que hay que tener cuidado.

Escalado lineal dinámico: es cuando el parámetro b varía con cada generación.

$$f_k' = a \cdot f_k + b_t$$

Una posible definición para b_t es darle el valor del menor fitness en la población actual t $b_t = -f_{min}$.

Truncación sigma: este método fue sugerido por Forrest con el propósito de mejorar el escalado lineal y poder trabajar con valores de fitness negativos e incorporar la información dependiente del problema dentro de la aplicación.

$$f_k' = f_k - (\bar{f} - c\sigma)$$

Donde c es un entero pequeño, σ es la desviación estándar de la población, y \bar{f} es el valor medio del fitness de la población. Los posibles valores de fitness escalado negativos son configurados a 0.

Escalado potencial : Este método fue propuesto por Gillies :

$$f_k' = f_k^\alpha$$

El valor de α depende del problema que se considere ($\alpha=1.005$). Se ha de indicar que la distancia entre los mejores cromosomas y los peores va aumentando conforme aumenta el valor de α . Cuando α se aproxima a 0 la distancia va disminuyendo y el muestreo se configura como una búsqueda aleatoria.. Cuando $\alpha > 1$, la distancia aumenta y el muestreo se concentra en los mejores individuos. Quizás el valor de α deba ser modificado a lo largo de una carrera de un algoritmo genético.

Otra posible definición sería:

$$f_k' = (a f_k + b)^\alpha$$

Normalización: Está dentro del escalado dinámico. Para problemas de maximización toma la siguiente forma:



$$f'_k = \frac{f_k - f_{min} + \gamma}{f_{max} - f_{min} + \gamma}$$

f_{min} y f_{max} son el peor y el mejor fitness de la generación actual.

γ es un número real positivo dentro del intervalo (0,1)

El propósito de utilizar esta trnasformación es:

1. Prevenir que la ecuación presente una división por 0.
2. Hacer posible el ajuste de la presión de selección, pasando de una selección proporcional al fitness a una selección puramente aleatoria.

Ranking: La idea es clara, ordenar la población de mejor individuo al peor y asignar una probabilidad de selección a cada cromosoma de acuerdo a su clasificación y no a su valor de fitness.

Sea p_k la probabilidad de selección para el k-esimo cromosoma en la población clasificada, la clasificación lineal toma la siguiente forma :

$$p_k = q - (k-1) \cdot r$$

Donde q es la probabilidad para el mejor cromosoma.

Si q_0 es la probabilidad para el peor cromosoma el parámetro r se puede determinar de la siguiente manera:

$$r = \frac{q - q_0}{pop_size - 1}$$

Los valores de fitness intermedios son disminuidos desde q hasta q_0 proporcionalmente a su clasificación. Cuando $q_0=0$ se tiene la mayor presión de selección

Michalewicz propuso el siguiente método de clasificación exponencial:

$$p_k = q(1 - q)^{k-1}$$

Un valor de q más elevado supone una mayor presión de selección.

1.2.4 Presiones de selección.

El proceso de evolución se puede dividir en tres categorías:

1. *Selección estabilizadora*, también conocida como selección normalizada tiende a eliminar cromosomas con valores extremos.
2. *Selección direccional*, tiene el efecto de aumentar o disminuir el valor medio de la población. La mayoría de los métodos de selección están basados en están basados en selección direccional.



3. *Selección disruptiva* tiende a eliminar cromosomas con valores de fitness moderados.

2. PROBLEMAS DE OPTIMIZACION CON RESTRICCIONES.

2.1 Optimización sin restricciones.

2.1.1 Minimización de la función de Ackley mediante la utilización de un AG.

La función de Ackley presenta la siguiente forma:

$$f(x_1, x_2) = -c_1 \cdot e^{\left(-c_2 \sqrt{\frac{1}{2}(x_1^2 + x_2^2)}\right)} - e^{\frac{1}{2}(\cos(c_3 x_1) + \cos(c_3 x_2))} + c_1 + e$$

$$-5 \leq x_1 \leq 5 \quad -5 \leq x_2 \leq 5$$

$$c_1 = 20 \quad c_2 = 0.2 \quad c_3 = 2\pi \quad e = 2.71282$$

Esta función presenta un mínimo global en(0,0) con un valor de 0.

Utilizaremos la siguiente implementación para el AG:

1. Codificación con números reales. El cromosoma consta de dos números reales. $\mathbf{v}=(x_1, x_2)$.
2. Cruce aritmético.
3. Mutación no uniforme.
4. Selección de los mejores *pop_size* .

Con respecto al punto 4 indicar que la siguiente generación se produce mediante la selección de los *pop_size* mejores individuos entre los padres y los hijos. Únicamente utilizamos el valor de fitness, para realizar la clasificación de los cromosomas.

Los parámetros del AG se configuraron de la siguiente manera:

pop_size=10
maxgen=1000
p_m=0.1
p_c=0.3

El algoritmo sería de esta forma:

ALGORITMO GENETICO

t=1

Generación de una población inicial aleatoriamente.

Evaluación de la población inicial

While (*t*<genmax)

t=t+1 ;

Operador de cruce aritmético sobre P, genera H1 hijos.

Operador de mutación no uniforme sobre P, genera H2 hijos.

Evaluación de los hijos generados .

Selección de pop_size mejores individuos entre P y H

end



Resultado :

En la generación 1000 se obtuvieron los siguientes cromosomas :

	X₁	X₂		X₁	X₂
v₁	-0.000002	-0.000000	v₆	-0.000002	-0.000000
v₂	-0.000002	-0.000000	v₇	-0.000002	-0.000000
v₃	-0.000002	-0.000000	v₈	-0.000002	-0.000000
v₄	-0.000002	-0.000000	v₉	-0.000002	-0.000000
v₅	-0.000002	-0.000000	v₁₀	-0.000002	-0.000000

El valor de fitness era: $f(x_1, x_2) = -0.005456$

2.2 Programación no lineal.(Optimización con restricciones)

Trata con el problema de obtener el óptimo de una función de coste sujeta a restricciones de igualdad y a desigualdades.

2.2.1 Tratamiento de las restricciones.

El problema central de la aplicación de los AG a la optimización con restricciones es como manejar dichas restricciones ya que los operadores que manipulan los cromosomas a menudo generan hijos no válidos.

Las técnicas actuales para resolver este problema se pueden clasificar en los siguientes tipos:

- Estrategias de rechazo.
- Estrategias de reparación.
- Estrategias de operadores genéticos modificados.
- Estrategias de penalización.

Estrategias de rechazo. Descartan aquellos cromosomas que no cumplen con las restricciones y que se van generando en el proceso de evolución.

- *Ventajas:* funciona razonablemente bien si el espacio de búsqueda válido es convexo.
- *Desventajas:* A menudo el sistema puede alcanzar más fácilmente el óptimo si se le permite cruzar una zona de puntos no válidos, sobre todo en espacio de búsqueda válidos no convexos.

Estrategias de reparación. Toman a un cromosoma que se haya generado y que no sea válido y mediante un algoritmo de reparación lo convierte en un cromosoma válido.

- *Desventajas:* Es dependiente del problema, es decir, para cada problema hay que generar un algoritmo de reparación distinto. En muchas ocasiones el encontrar este algoritmo es casi tan difícil como resolver el problema original.



Estrategias de operadores genéticos modificados. Consiste en generar una representación de los cromosomas y unos operadores genéticos que durante el proceso de evolución siempre trabajen con soluciones válidas. Michalewicz consideraba que en muchas ocasiones estos sistemas son mucho más realizables que los basados en estrategias de penalización.

Estrategias de penalización. Las tres estrategias anteriores únicamente consideran movimientos dentro de la región de búsqueda válida. Pero en muchas ocasiones el buscar el óptimo permitiendo movimientos dentro de las zonas de búsqueda no válidas produce una optimización mucho más rápida y unas soluciones mucho mejores, que si únicamente se permite el movimiento dentro de la zona de búsqueda válida.

Las estrategias de penalización trabajan también con puntos en la región de búsqueda no válida. Esencialmente esta estrategia lo que pretende es convertir un problema con restricciones a otro sin restricciones mediante la penalización de soluciones no válidas, de hecho un término de penalización se añade a la función objetivo para cualquier violación de las restricciones.

$$eval(x)=f(x)+p(x)$$

En general el espacio de solución contiene dos partes : un área de puntos válidos y un área de puntos no válidos.

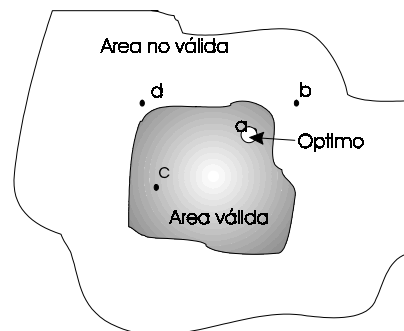


Figura 2: Espacio de soluciones

De la Figura 3 se puede ver como la solución no válida b está mucho más cerca del óptimo a que la solución válida c y que la solución no válida d. Podría esperarse dar una penalización menor a b que a d, aunque b este más alejada de la región válida que d. También podríamos pensar que b contiene más información sobre el óptimo que c, a pesar de ser una solución no válida. Sin embargo en la mayoría de las ocasiones no se conoce el óptimo y no se pueden realizar este tipo de juicios. La elección de la función de penalización $p(x)$ es fundamental para realizar una estrategia de búsqueda adecuada.

Si se tiene que:

$$eval(x)=f(x)+p(x)$$

Se requiere que:



$$\begin{aligned}
 p(x) &= 0 \quad \text{si } x \text{ es válido} \\
 p(x) &< 0 \quad \text{en otro caso (maximización)} \\
 p(x) &> 0 \quad \text{en otro caso (minimización)}
 \end{aligned}$$

Además se requiere que:

$$|p(x)|_{\max} \leq |f(x)|_{\min}$$

para evitar tener valores de fitness negativos.

Otra forma de realizar la penalización es multiplicando por la función de coste:

$$eval(x) = f(x) \cdot p(x)$$

En este caso se requiere que:

$$\begin{aligned}
 p(x) &= 1 \quad \text{si } x \text{ es válido} \\
 0 \leq p(x) &< 1 \quad \text{en otro caso (maximización)} \\
 p(x) &> 1 \quad \text{en otro caso (minimización)}
 \end{aligned}$$

Clasificación de la funciones de penalización

Se han propuesto muchas y muy variadas funciones de penalización, que se pueden agrupar en dos grandes grupos:

- Penalización constante
- Penalización variable.

La *penalización variable* es mucho más efectiva que la constante, tiene dos componentes :

1. Razón de variación de la penalización.

Este factor puede ser ajustado de acuerdo a :

- *El grado de violación de la restricción.* Incrementa la presión de penalización cuando la violación es más importante, forma la clase conocida como penalización estática.
- *El número de iteración del algoritmo.* Aumenta la presión de penalización conforme aumentan las iteraciones del algoritmo.

2. Cantidad de penalización para la violación de las restricciones.

Esencialmente la función de penalización es una función de la distancia a la zona de búsqueda válida. Esto puede ser dado de tres posibles maneras:

- La función de la distancia absoluta de una única solución no válida.
- La función de la distancia relativa de todas las soluciones no válidas en la población actual.
- La función del término de penalización adaptativo.

Las funciones de penalización también se pueden distinguir como :

- Dependientes del problema.
- Independientes del problema.



Veamos algunas funciones de penalización

Método de Homaifar, Qi y Lai. Se consideraba el siguiente problema de programación no lineal:

$$\begin{aligned} &\min(f(\mathbf{x})) \\ &\text{sujeta a } g_i(\mathbf{x}) \geq 0 \quad i=1,2,\dots,m \end{aligned}$$

Se trabajo con la siguiente función de evaluación

$$eval(\mathbf{x})=f(\mathbf{x})+p(\mathbf{x})$$

La función de penalización es construida con dos componentes de la siguiente manera:

$$p(x) = \begin{cases} 0 & \text{si } x \text{ es válido} \\ \sum_{i=1}^m r_i g_i^2(x) & \text{en otro caso} \end{cases}$$

Donde r_i es el coeficiente de penalización variable para la i -ésima restricción. Para cada restricción existen varios niveles de violación. Dependiendo del nivel de violación, r_i varia. Sin embargo la determinación del nivel de violación para cada restricción y elegir un valor de r_i adecuado no es tarea fácil y depende del problema sobre el que se aplique el algoritmo.

Michalewicz indico que la calidad de las soluciones depende de los valores de estos coeficientes de penalización. Si los coeficientes de penalización son moderados quizás converja a una solución que no es válida, por otro lado si son muy elevados estaríamos realizando una estrategia de rechazo.

Método de Michalewicz y Attia:

Consideraban el siguiente problema de programación no lineal :

$$\begin{aligned} &\min(f(\mathbf{x})) \\ &\text{sujeta a } g_i(\mathbf{x}) \geq 0 \quad i=1,2,\dots,m_1 \\ &h_i(\mathbf{x})=0 \quad i=m_1+1,\dots,m \quad m=m_1+m_2 \end{aligned}$$

Toma la forma aditiva de la función de penalización:

$$eval(x)=f(x)+p(\tau,x)$$

donde:

$$p(\tau, x) = \frac{1}{2\tau} \sum_{i \in A} d_i^2(x)$$

Donde A es el conjunto de restricciones activas, que consiste de todas las ecuaciones no lineales y de todas las desigualdades no lineales violadas.



Una restricción $g_i(x)$ es violada en el punto x si y solo si $g_i(x) > \delta$ ($i=1, \dots, m_1$), donde δ es un parámetro para decidir si una restricción está activa.

τ es la componente de penalización variable, llamada temperatura.

El termino de penalización para una única restricción $d_i(x)$ esta dada por:

$$d_i(x) = \begin{cases} \max\{0, g_i(x)\} & \text{para } 1 \leq i \leq m_1 \\ |h_i(x)| & \text{para } m_1 + 1 \leq i \leq m \end{cases}$$

La temperatura τ comienza con un valor inicial τ_0 y finaliza a una temperatura τ_f de enfriamiento, la temperatura va disminuyendo de acuerdo a algún esquema de enfriamiento.

Ellos construyen un sistema conocido como GENECOP II

