

03063

11



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

UNIDAD ACADÉMICA DE LOS CICLOS  
PROFESIONALES Y DE POSGRADO

Métodos de optimización con restricciones  
usando Algoritmos Genéticos y Funciones de  
Penalización. Análisis comparativo.

T E S I S  
QUE PARA OBTENER EL GRADO DE:  
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN  
P R E S E N T A:  
JUAN JESÚS GUTIÉRREZ GARCÍA

916762

Director de tesis  
Dr. Ángel Fernando Kuri Morales

Junio 2001



Universidad Nacional  
Autónoma de México



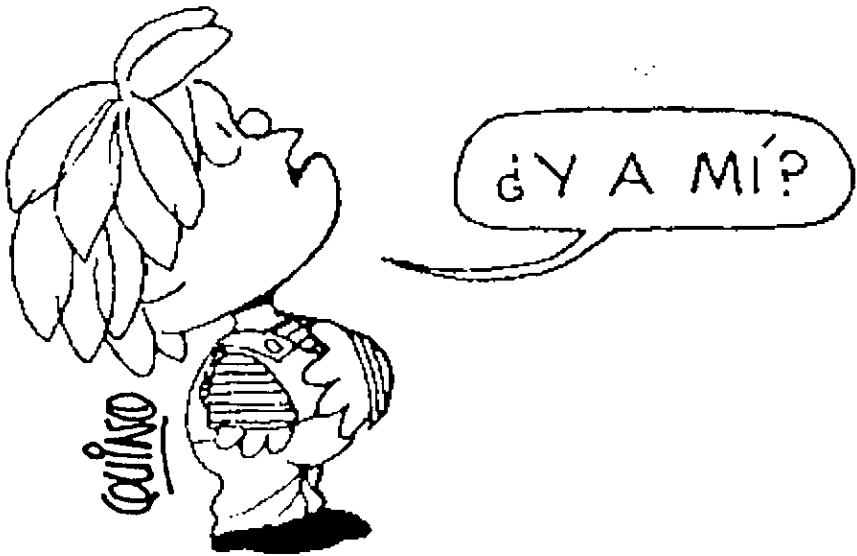
**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**A Dios,**  
*A mi ESPOSA,*  
*A mi MADRE y*  
*A la memoria de*  
*mi PADRE*



## Agradecimientos

*Escribir estas líneas me provoca sentimientos encontrados: Gusto por dar a quienes me han ayudado un agradecimiento, pero también desilusión porque será imposible mencionarlos a todos.*

*Agradezco mucho al Dr. Ángel Kuri por enseñarme que no existe motivo para no titularse y por el apoyo que me ha dado para terminar este trabajo.*

*Agradezco a mi esposa por su paciencia en todos esos días de trabajo y por el apoyo que siempre me proporciona.*

*Agradezco a mi mamá su insistencia para que este trabajo llegara a término y por todo lo que siempre me ha proporcionado.*

*Agradezco a mis hermanos, familiares, amigos y compañeros de trabajo tanto su apoyo como las porras para terminar.*

*Su Amigo de siempre  
Juan Jesús Gutiérrez García*

## **Contenido**

<u>Métodos de optimización con restricciones usando algoritmos genéticos y funciones de penalización. Análisis comparativo</u> .....	1
<u>Contenido</u> .....	i
<u>Capítulo 1. Introducción</u> .....	1
<u>Capítulo 2. Algoritmos Genéticos</u> .....	3
<u>2.1 Descripción</u> .....	3
<u>2.2 Elementos de los Algoritmos Genéticos</u> .....	5
<u>2.3 Esquemas</u> .....	11
<u>Capítulo 3. Métodos de Optimización con Restricciones</u> .....	14
<u>3.1 Optimización sin restricciones</u> .....	14
<u>3.2 Optimización con restricciones</u> .....	16
<u>Capítulo 4. Métodos Comparados</u> .....	21
<u>4.1 Método de Homaifar et al</u> .....	22
<u>4.2 Método de Joines y Houck</u> .....	25
<u>4.3 Método de Schoenauer y Xanthakis</u> .....	27
<u>4.4 Método de Powell y Skolnick</u> .....	28
<u>4.5 Método de Kuri</u> .....	28
<u>Capítulo 5. Análisis comparativo</u> .....	30
<u>5.1 Sustento estadístico</u> .....	30
<u>5.2 Diseño y desarrollo de los experimentos</u> .....	34
<u>5.3 Resultados</u> .....	35
<u>5.4 Comentarios a los resultados</u> .....	37
<u>Capítulo 6. Conclusiones</u> .....	42
<u>6.1 Trabajos Futuros</u> .....	42

<u>Apéndice I. Descripción de las funciones de prueba</u> .....	44
<u>I.1 Funciones</u> .....	45
<u>I.2 Gráficas</u> .....	52
<u>Apéndice II. Descripción del software utilizado</u> .....	56
<u>Apéndice III. Aptitud vs. Generación</u> .....	59
<u>Bibliografía y Referencias</u> .....	60

## **Capítulo 1. Introducción.**

Los problemas de optimización han gozado de gran interés porque representan situaciones del mundo real y el caso de la optimización restringida tiene muchas aplicaciones. En los libros que tratan el tema los ejercicios propuestos son, en muchos casos, problemas reales [11]. Tal ha sido el interés por éstos que al iniciar el presente trabajo se ha podido apreciar cómo las funciones con las cuales se prueban los métodos de optimización se repiten frecuentemente de un artículo a otro [19],[7],[25] y en su mayoría están tomados de un libro de optimización restringida.

Es la misma naturaleza que nos ha propuesto estos complicados problemas la que nos ha proporcionado un modelo para empezar a resolverlos: los Algoritmos Genéticos (AG). La gran diversidad de una población de seres vivos, le ayuda a sobrevivir frente a cambios del medio ambiente y es mediante el proceso de evolución que estas poblaciones se adaptan a los cambios. Holland en 1975 propone utilizar un método que simula la evolución biológica y en 1989 Goldberg popularizó éste llamándole Algoritmo Genético Simple (AG)[7].

Los AG, han mostrado ser particularmente efectivos en problemas de optimización difíciles [9],[21],[20], con un espacio de solución no convexo, problemas discontinuos, continuos, discretos y de variables de cualquier tipo.

Estas son las razones por las cuales se busca aplicar el AG a la solución de problemas restringidos. Los algoritmos genéticos pueden ser utilizados directamente para resolver problemas de optimización no restringidos. En el caso de optimización restringida se procede de dos formas: se modifican los operadores del AG o se modifica el problema para transformarlo en un problema no restringido y así aplicar el AG. Con este segundo enfoque se trabaja esta tesis.

A mediados de la década de 1990 se publicaron varios artículos en los que se combinan una función de penalización con algunos coeficientes y los AGs. La función objetivo es penalizada si alguna de las restricciones no se cumple. El trabajo de esta tesis consiste en realizar un análisis estadístico para comparar algunos de estos métodos y concluir cuál de ellos es mejor.

La inquietud de realizar este análisis surge de [2] donde se describen varios métodos que utilizan AGs para resolver el problema de la optimización restringida, pero ¿cuál de estos métodos se debe utilizar? ¿hay alguno que sea mejor que los otros?

Estas preguntas continúan sin respuesta absoluta aún después de terminado el trabajo. La razón es que existe una gran cantidad de métodos descritos como para realizar el análisis de todos ellos (se propone como trabajos futuros en las conclusiones hacer el análisis sobre otros). Por tal razón sólo se utilizaron métodos con una función de penalización. Pero aun así se han encontrado

resultados interesantes y sorprendentes que serán de utilidad para trabajos posteriores.

El análisis resultó favorable porque uno de los métodos muestra un mejor desempeño, por un orden de diez, del resto. Estos resultados no se aplican a todos los posibles problemas de optimización, se han seleccionado 25. Pero un comportamiento similar se espera para la mayor parte de los problemas existentes dado que los seleccionados no tienen características en común.

En la primera parte del trabajo se describe brevemente algunos de los conceptos de los algoritmos genéticos, sus operadores y el teorema del esquema. Posteriormente se describen algunos de los métodos clásicos con los cuales se ha tratado de resolver el problema de optimización restringida.

En el capítulo 3 se describe cada uno de los métodos a analizar. Se han seleccionado los que utilizan funciones de penalización. Estos métodos también se describen en el libro de Z. Michalewicz [25] y se realiza un análisis comparativo menos riguroso que el mostrado en esta tesis (prueba sólo con cinco funciones y realiza 10 ejecuciones de cada método).

El método de Homaifar (que ha sido el que tiene un peor desempeño en este análisis) ha presentado algunos problemas que se describen en la sección 4.1. Algunos de los problemas presentados ya se han reportado en [24].

El capítulo cinco describe las técnicas estadísticas que se utilizaron, el número de experimentos realizados y los resultados obtenidos. Y en el capítulo 6 se construye una tabla que compara cada uno de los métodos. El mejor de ellos es 1.61 veces en 2 de cada 3 casos que el método de Homaifar.

Finalmente se presentan tres apéndices que describen: las 25 funciones de prueba que se utilizaron para el análisis con las gráficas de algunas de ellas, el software construido para implementar cada uno de los métodos estudiados y una descripción del comportamiento de las gráficas del mejor valor para la aptitud vs. generación.



## **Capítulo 2. Algoritmos Genéticos.**

En el presente capítulo se expondrán los conceptos básicos de los algoritmos genéticos. No se trata de hacer un estudio detallado de cada uno de sus aspectos; por el contrario, sólo se comentan en forma breve para que el lector esté familiarizado con los términos que se utilizan en el resto de la tesis y se hace referencia a libros que comentan ampliamente cada tema. La mayor parte de la información presentada está tomada de [19].

### **2.1 Descripción.**

En una ciencia nueva, como lo es la computación, no existe consenso por las definiciones, por este motivo para quedar en claro sobre los algoritmos genéticos (AG) se presentan a continuación tres de ellas:

*"Genetic Algorithms (GAs) are computer-oriented procedures which attempt to characterize the essentials of system by partially simulating a process of natural selection" [19].*

*"Genetic algorithms are search algorithms based on the mechanics of natural selection and natural genetics. They combine survival of the fittest among string structures with a structures yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. In every generation, a new set of artificial creatures (strings) is created using bits and pieces of the fittest of the old; an occasional new part is tried for good measure" [7].*

*" ... For small spaces, classical exhaustive methods usually suffice; for large spaces special artificial intelligence techniques must be employed. Genetic Algorithms (GAs) are among such techniques; they are stochastic algorithms whose search methods model some natural phenomena: genetic inheritance and Darwinian strife for survival" [25].*

Para poder emplear un AG en un problema de búsqueda es necesario que cualquier posible solución se pueda codificar. Este es el primer paso al implantar un AG; para hacerlo es necesario conocer el "tamaño" del espacio de búsqueda (en caso de ser infinito se tendrá que hacer una aproximación) y obtener una codificación de las posibles soluciones.

Posteriormente se crea una población inicial. Esto es un conjunto de posibles soluciones, que pueden ser generadas aleatoriamente o no. De esta primera población se selecciona un conjunto de individuos (soluciones posibles) que sean más aptos que el resto. La aptitud de un individuo se mide por una función (que debe estar definida para todo el espacio de búsqueda) que tomará valores mayores para los individuos que cumplen con las características deseadas y menores para los que carecen de éstas. La acción de calificar a cada individuo

con un número se conoce como evaluación. Cada individuo de la población debe ser evaluado para conocer qué tan apto es.

Después de realizar la evaluación se selecciona a un conjunto de individuos que serán cruzados (este es un método para a partir de dos soluciones obtener otra) y se inducen mutaciones (cambios aleatorios en la codificación de cada individuo) para que formen la nueva generación.

Se evalúa la nueva generación y el proceso se repita hasta cumplir la condición de término (puede ser que se ha encontrado al individuo buscado o que se está muy cerca de él).

Un AG converge a la solución cuando entre los elementos de la población aparecen individuos que son la solución del problema o que están "muy cerca" de ésta.

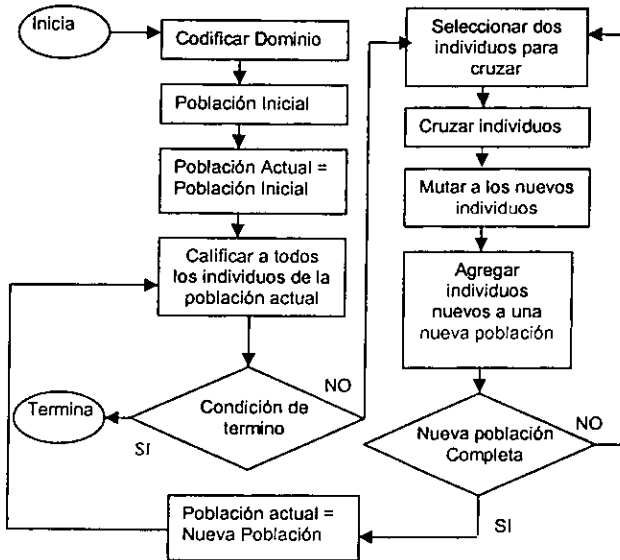


Figura 1. Algoritmo Genético Básico. Cada vez que se genera una nueva población se dice que ha pasado una generación

La Figura 1 muestra un diagrama con el ciclo básico de un algoritmo genético. El número de generaciones que se efectuarán (esto es cada vez que la población actual se convierte en la nueva población) queda determinado por las condiciones de término. Estas condiciones pueden ser: que ya se haya encontrado el valor buscado, que se esté a una distancia  $\epsilon$  del valor deseado, que los cambios en la población de una generación a otra sean mínimos, entre otras.

Cuando la selección es proporcional y la cruza en un punto (posteriormente se explican con detalle estos conceptos) al algoritmo genético se le conoce como "simple" (AGS)[7].

Los AGs pertenecen a un grupo más grande de técnicas conocidas como computación evolutiva entre estas se encuentra la Programación Evolutiva, las estrategias evolutivas y los AG. Algunas de las características de los AGs, que les ha dado su popularidad, son:

- Son muy flexibles. Es decir, pueden adoptar con facilidad nuevas ideas, generales o específicas, que surjan dentro del campo de la computación evolutiva. Además se pueden combinar fácilmente con otros paradigmas y enfoques, aunque no tengan ninguna relación con la computación evolutiva.
- De entre todos los paradigmas de la computación evolutiva son los que menos conocimiento específico necesitan para su funcionamiento, y en consecuencia, los más versátiles. Y pueden incorporar conocimiento específico con poco esfuerzo adicional.
- Son fáciles de implantar en computadoras con capacidades medias, proporcionando resultados aceptables (en cuanto a precisión y recursos empleados) para una gran cantidad de problemas difícilmente resolubles por otros métodos.
- Existe una gran cantidad de ensayos empíricos que proporcionan operadores, parámetros e implantaciones específicas para una amplia gama de problemas.

## *2.2 Elementos de los Algoritmos Genéticos.*

En la sección anterior se habló en forma general de los AGs. En este apartado se describen cada uno de sus elementos. Cabe mencionar que existen variaciones de dichos elementos y no será posible explicar todos ellos. En la bibliografía se cita varios libros que tratan estos temas.

Aún cuando los algoritmos genéticos se pueden ocupar en muchos y muy diversos problemas en esta tesis sólo se tratará lo referente a la solución de maximizar o minimizar funciones de  $\mathfrak{R}^n \rightarrow \mathfrak{R}$ . Además, minimizar una función es equivalente a maximizar el negativo de esta función. Por tal razón en el texto únicamente se utiliza la segunda de estas palabras y se sobrentiende que es posible resolver ambos problemas indistintamente.

### 2.2.1 Codificación.

Cuando se tiene un problema de búsqueda existe un conjunto de posibles soluciones de las cuales se distingue una de ellas por cumplir con cierta propiedad. Algunos ejemplos son: maximizar una función (que es el problema que

se estudia en el resto de la tesis), hallar la solución a un conjunto de ecuaciones no lineales, encontrar la estrategia óptima en un juego y otros más. Para algunos métodos es necesario que la representación de esta solución destaque la característica que se desea encontrar. En el caso de los AG esto no es necesario.

Se entenderá por codificar, en esta tesis, el construir una función uno a uno y con dominio en el espacio de soluciones del problema e imagen en cadenas de símbolos ya definidos. Esto es, cada posible solución está asociada con una cadena. A éstas se les conoce como genomas o cromosomas, cuando están formados sólo por los símbolos 0,1 reciben el nombre de cadenas de *bits*.

Esta función de codificación hace que una posible solución  $x$  sea representada como una cadena de símbolos  $s$ . Entonces se le llama fenotipo a  $x$  y genotipo a  $s$ .

Habrás casos, como el problema de maximizar una función de  $\mathcal{R}^n \rightarrow \mathcal{R}$ , en los cuales no se podrá encontrar la función de codificación y será suficiente con tener una aproximación. Esto significa que una cadena representa a un conjunto de soluciones y la respuesta será encontrar la cadena que representa al conjunto que contiene la solución o una muy "similar".

Existen muchas formas de codificar el espacio de soluciones, sólo se tratarán las codificaciones de números reales en binario pesado y Gray en punto fijo. Pero también se puede codificar con punto flotante [19] o utilizando cadenas con otros símbolos [25].

En la representación en punto fijo se tiene una pareja de números enteros y el número real se forma concatenándolos con un punto en medio de ellos. En cambio, en punto flotante, el segundo número entero determina la posición del punto decimal en el otro. Ejemplos de algunas parejas de enteros y su significado se muestran en la Tabla 1.

Tabla 1. Ejemplos de punto fijo y punto flotante

Pareja de números		Valor en punto fijo	Valor en punto flotante
1	1	1.1	$0.1 \times 10^1 = 1$
124	3	124.3	$0.124 \times 10^3 = 124$
4536	12	4536.12	$0.4536 \times 10^{12} = 4536 \times 10^8$

Para poder representar un número real, usando la codificación binaria pesada y punto fijo, se puede convertir la parte entera en una serie de ceros y unos que representen los coeficientes de una suma de potencias en base 2

$$ent(x) = \sum_{i=0}^n c_i 2^i \text{ donde } c_i \in \{0,1\} \quad (1)$$

La parte fraccionaria se representa con el conjunto de  $d$ 's que hagan mínima la siguiente resta:

$$frac(x) - \sum_{i=1}^m \frac{d_i}{2^i} \text{ donde } d_i \in \{0,1\} \quad (2)$$

$frac(x)$  representa la parte fraccional del número real  $x$ .

Esto implica que una infinidad de números tendrán la misma representación. Esta es la causa de que no se tenga una solución exacta sino aproximada y mientras mayor sea el número  $m$  menor será el error por la representación.

De las ecuaciones (1) y (2) se tiene una cadena de codificación con longitud  $m+n+1$ , el uno se agrega por el signo de todo el número. Al utilizar esta codificación es necesario tener  $n$  dígitos para la parte entera y  $m$  dígitos para la parte fraccionaria.

En el caso de la codificación Gray se utiliza la cadena de ceros y unos en binario pesado (que se identificarán con  $p_i$  y la cadena en Gray con  $g_i$ ) y se construye con la siguiente regla: el primero de ellos (el bit de la izquierda) se copia (así se obtiene  $g_0$ ) y después se aplica un XOR entre el  $g_0$  y  $p_1$  para obtener  $g_1$  en general  $g_i = g_{i-1} XOR p_i$ .

En la siguiente tabla se muestran algunos ejemplos de codificación en binario pesado y en Gray, con 8 dígitos enteros y 8 decimales

Tabla 2. Codificación Gray y binario pesado. El 5.1 en ambas codificaciones es un número periódico y sólo se puede representar una aproximación.

Número	Gray	binario pesado
1	00000001.00000000	00000001.00000000
2	00000010.00000000	00000011.00000000
3	00000011.00000000	00000010.00000000
45.25	00111011.11110000	00101101.01000000
5.1	00000111.10010101	00000101.00011001

Pero en la mayor parte de los casos no se codifica un valor real, sino un vector y se concatenan cada una de las cadenas que representan las coordenadas del vector, así el tamaño del cromosoma estará dado por:

$$Dim(v)(m+n+1) \quad (3)$$

Donde  $Dim(v)$  es la dimensión del vector que se representa,  $n$  el número de dígitos binarios que se utilizan para la parte entera de cada coordenada y  $m$  para la parte decimal.

## 2.2.2 Evaluación.

Para evaluar a los individuos de la población se necesita una función que les asigne un valor numérico a cada uno. Entre mayor sea este valor se dice que el individuo es más apto y si el valor es pequeño entonces es poco apto. A esta medida de bondad se le conoce como función de aptitud.

Es necesario tomar en cuenta que la función de aptitud no debe tomar valores negativos cuando la selección es proporcional. Si el problema es maximizar  $f(x)$  se deberá de evaluar las siguientes:

$$f'_i = f_i + \frac{1}{N} \sum_i |f_i| + |\min(f_i)| \quad \forall i \quad (4)$$

$$F_i = \frac{f'_i}{\sum_{i=1}^N f'_i} \quad (5)$$

Aquí la variable  $i$  representa a cada individuo de la población,  $N$  es el total de los individuos y  $F$  será la función de aptitud normalizada [19].

La función de aptitud evalúa el cromosoma. Sólo es necesario poner atención en el tipo de codificación que se utiliza para la conversión a número real.

### 2.2.3 Selección.

La selección de los mejores individuos puede ser probabilística (por ejemplo proporcional, torneo) o determinista (por ejemplo Vasconcelos, Nietzsche) [19]. Aquí sólo se describe un método proporcional (ruleta) y uno determinista (Vasconcelos).

Una vez que se tiene la aptitud normalizada usando una de las técnicas de selección proporcional se puede construir una ruleta que tiene un área equivalente al valor de la aptitud asignado a cada individuo; ver la Figura 2.

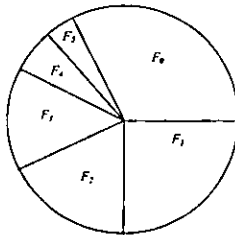


Figura 2. Ruleta de valores proporcionales para cada uno de los individuos dependiendo de su aptitud.

Para determinar cuáles parejas de individuos serán seleccionadas para cruzarse se selecciona un par de puntos del círculo al azar. Es claro que tendrán mayores probabilidades de ser seleccionados aquellos con un área mayor. Esto significa que obtuvieron un valor de aptitud mayor, que corresponden a los que tienen la mayor adaptabilidad.

En cambio, el modelo de Vasconcelos, siempre construye las parejas para realizar la cruce seleccionando al mejor de los individuos y el peor de ellos; posteriormente al segundo mejor con el penúltimo. En el ejemplo de la Figura 2 se ha marcado cada "rebanada" del círculo con letras  $F$ . Ordenadas de mayor ( $F_6$ ) a menor ( $F_5$ ). Al

utilizar el modelo de Vasconcelos para estos valores se tendrán las siguientes parejas:  $(F_0, F_3)$ ,  $(F_1, F_4)$  y  $(F_2, F_3)$ .

A las parejas seleccionadas se aplica el procedimiento de cruce y se generan los descendientes que estarán en la siguiente generación.

La selección es el operador que permite al AG genético aprender qué individuos son más aptos y sus características deben ser conservadas.

#### 2.2.4 Elitismo.

El elitismo consiste en conservar 1 o más individuos de la generación actual para la siguiente. Es claro que los individuos que pasan de una generación a otra son los que tienen mayor valor de la función de aptitud y estos pueden participar o no en el cruzamiento para generar nuevos individuos. Los programas que se realizaron como parte de esta tesis utilizan la segunda opción.

También se ha implementado el elitismo total [19]. Si se tiene una población con  $N$  individuos entonces se crea una nueva población con todos los individuos de la generación anterior y los de la actual (ésta tendrá más de  $N$  individuos y si el número de individuos no cambia entre generaciones será  $2N$ ), se ordenan de acuerdo a su aptitud y se construye la siguiente generación con los mejores  $N$  individuos de ambas (este método se puede ver como una variante de la selección  $(\mu+\lambda)$  que se usan en Estrategias Evolutivas conocida como "*Selective Breeding*").

La finalidad del elitismo es aumentar la convergencia del AG a una solución conservando toda la información de los mejores individuos de una generación a otra. Si no hay elitismo se conserva sólo parte de información de los mejores individuos (en sus descendientes) en la siguiente generación.

#### 2.2.5 Cruzamiento.

Es posible hacer cruzamientos de individuos tanto en el fenotipo como en el genotipo. Por ejemplo, si el espacio de búsqueda consiste de números reales, se puede cruzar a dos individuos tomando su valor medio (fenotipo) o intercambiando sus bits (genotipo). Es claro que un cruzamiento en el fenotipo corresponde a uno en el genotipo e inversamente, y afectar al genotipo sólo facilita su implantación. Para este trabajo sólo se utilizaron los siguientes tipos de cruzamiento:

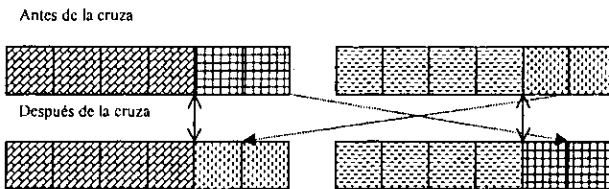


Figura 3. Cruza en un punto

## a) 1 - punto

Consiste en seleccionar una posición válida al azar del cromosoma e intercambiar las cadenas restantes como se muestra en la Figura 3. Las flechas verticales indican que la posición seleccionada ha sido la 4. Posiciones inválidas son 0 y 6 pues no generarían ningún cambio.

## b) Dos Puntos (anular)

Consiste en seleccionar dos posiciones en el cromosoma e intercambiar las cadenas como se muestra en la Figura 4. Se han seleccionado las posiciones 3 y 5

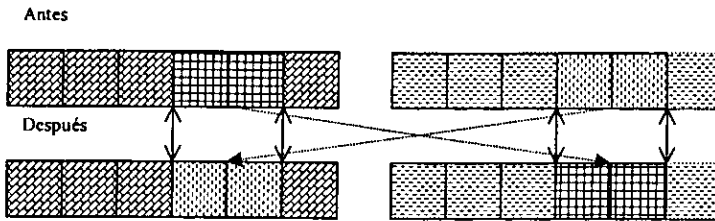


Figura 4. Cruzamiento en dos puntos

## c) Uniforme

Se eligen  $n$  números; aleatorios para cada posición del cromosoma se compara el número aleatorio (que está entre 0 y 1) con 0.5. Si es mayor entonces se toma el cromosoma del padre y si es menor se toma el de la madre. De forma contraria para el otro hijo. Esto se ilustra en la Figura 5. En la parte derecha se han incluido los números aleatorios y según éstos se ha realizado el cambio del bit o no. Se ha ilustrado esto con letras y números que corresponden a los bits que estarían en esas posiciones

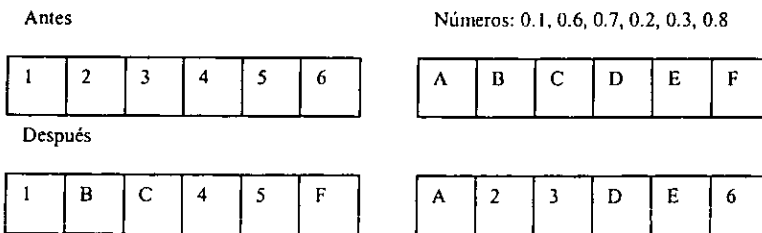


Figura 5. Cruzamiento uniforme

Uno de los parámetros del algoritmo genético es la probabilidad de que dos parejas seleccionadas se crucen. Está probabilidad por lo regular es grande (entre 0.7 y 0.9) porque se desea explotar lo más posible la información ya obtenida en generaciones anteriores. Si las parejas se cruzan los nuevos individuos son candidatos a ser agregados a la nueva población. De lo contrario, se puede elegir



entre proponer como nuevos individuos a los seleccionados o seleccionar otra pareja y probar cruzándolos nuevamente. En este trabajo se optó por el primer método.

El cruzamiento combina las características de los individuos para que a partir de la información que hay en su cromosoma se generen nuevos individuos con algunas de las características de los padres.

### 2.2.6 Mutación.

La mutación es el operador genético utilizado para explorar nuevas cromosomas, porque los derivados de la cruce contienen segmentos de cadenas que ya se habían evaluado con anterioridad, pero la mutación agrega segmentos de cadenas nuevas. Consiste en seleccionar un número aleatorio entre cero y uno para cada uno de los símbolos del cromosoma. Si el número es mayor que la probabilidad de mutación se complementa el *bit* en esa posición.

La probabilidad de mutación está frecuentemente en el intervalo de 0.01 a 0.005.

De lo expuesto hasta este momento se tiene que un algoritmo genético necesita de los siguientes parámetros para poder definirse con precisión:

- Probabilidad de Mutación
- Probabilidad de Cruza
- Criterio para detener su ejecución
- Tamaño de la población

No existen valores precisos para éstos y los rangos que se han obtenido son empíricos. Pero hay trabajos en los cuales algunos se han codificado dentro del mismo cromosoma para que cambien con cada generación del algoritmo genético. (ver por ejemplo [10] [4]).

### 2.3 Esquemas.

Los AG tienen fundamentos teóricos en, al menos, dos modelos: Los esquemas y las cadenas de Markov. A continuación se describe el teorema del esquema [12]. Para obtener información sobre las cadenas de Markov se puede consultar [19].

Para definir los esquemas se utiliza un conjunto de tres símbolos {0,1,\*}. El símbolo '\*' se conoce como no importa (*don't care*) y cada esquema representa un conjunto de cadenas de *bits* donde el '\*' representa a un 1 o un 0. Por ejemplo, 1\*0 representa a las cadenas 101 y 110. Se puede ver que mediante este concepto, se están definiendo regiones en el dominio del problema. Así en una representación binaria de longitud  $L$ , cada individuo pertenece a  $2^l$  esquemas o regiones del espacio de búsqueda. También se puede decir que un individuo es representante del esquema  $H$  si este pertenece a la región que forma  $H$  en el dominio.

Dado un esquema  $H$  su orden se define como el número de posiciones con valor explícito y se denota por  $o(H)$  y la distancia entre la primera y la última posición explícita se denomina longitud de definición y se denota con  $\delta(H)$ . En la Tabla 3 se muestran algunos ejemplos

Tabla 3. Esquemas con su orden y longitud

Esquema	$o(H)$	$\delta(H)$
1 0 * * * 1 * *	3	6 - 1 = 5
0 1 * 0 1 * 0 1 *	6	8 - 1 = 7
* 0 1 * * 1 * *	3	6 - 2 = 4
0 1 1 * 1 * *	4	5 - 1 = 4

Ahora se define la función  $m(H,t)$  que expresa el número de instancias de un esquema  $H$  en la generación  $t$  en una población  $A(t)$  con  $N$  individuos.

Si se considera una selección de ruleta la probabilidad de que el  $i$ -ésimo representante del esquema  $H$  sea seleccionado está dada por

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (6)$$

y la probabilidad de seleccionar a algún representante del esquema  $H$  será la suma de las probabilidades de cada representante del esquema en esa población

$$p_H = \frac{\sum_{j=1}^{m(H,t)} f_j}{\sum_{k=1}^N f_k} \quad (7)$$

la suma de las funciones de aptitud de cada uno de los representantes (numerador en la ecuación (7)) se puede escribir como el siguiente producto:

$$\sum_{j=1}^{m(H,t)} f_j = \bar{f}(H)m(H,t) \quad (8)$$

donde  $\bar{f}(H)$  es el valor de adaptación promedio de las instancias del esquema  $H$ .

Ahora si  $\bar{f}$  es el promedio de la adaptación de toda la población se cumple que

$$m(H,t+1) = m(H,t) \frac{\bar{f}(H)}{\bar{f}} \quad (9)$$

De aquí se aprecia que la selección provoca que instancias de un esquema en particular aumenten con la relación entre la aptitud promedio de ese esquema y la aptitud promedio de toda la población

Al considerar el efecto de la cruce en un punto junto con la selección se puede incluir un factor más en (9) (donde  $p_c$  es la probabilidad de cruce)

$$m(H,t+1) \geq m(H,t) \frac{\bar{f}(H)}{\bar{f}} \left[ 1 - p_c \frac{\delta(H)}{l-1} \right] \quad (10)$$

Este factor es el producto de la probabilidad de cruzar los individuos multiplicado por la probabilidad de que el esquema no sobreviva. Cambiará si se realiza una cruce en dos puntos o uniforme. El cálculo de éstos se puede consultar en [19]. Se trata de una desigualdad porque no se ha tomado en cuenta los nuevos esquemas, esto es, que dos individuos que no son representantes del esquema  $H$  al cruzarse generen a uno o más que sí pertenezcan a la región generada por  $H$ .

La ecuación (10) necesita contemplar además el operador de mutación que permitirá a un representante de un esquema continuar como tal si no se altera ninguno de sus bits explícitos (si se altera un bit que en el esquema no es explícito el nuevo individuo sigue siendo representante de éste). La probabilidad de que un esquema  $H$  no sea alterado está dada por:

$$(1 - p_m)^{o(H)} \quad (11)$$

en caso de que  $p_m$  sea pequeña ( $p_m \ll 1$ ) la expresión (11) se puede aproximar por  $1 - o(H)p_m$  y se tendrá:

$$m(H, t+1) \geq m(H, t) \frac{\bar{f}(H)}{\bar{f}} \left[ 1 - p_c \frac{\delta(H)}{l-1} - o(H)p_m \right] \quad (12)$$

La ecuación (12) es el llamado "teorema del esquema". Y expresa que el número de representantes de un esquema mejor calificado que el resto aumentará con cada generación.

Sea  $H$  un esquema en particular con una calificación media superior al promedio de la población ( $\bar{f} + c\bar{f}$ ) entonces el teorema del esquema se puede re-escribir (usando sólo la selección (9)) como sigue

$$m(H, t+1) = m(H, t) \frac{\bar{f} + c\bar{f}}{\bar{f}} = (1 + c)m(H, t) \quad (13)$$

donde  $c > 0$ . Si se supone además que el valor de  $c$  es estacionario (no varía entre generaciones) se obtiene

$$m(H, t) = m(H, 0)(1 + c)^t \quad (14)$$

Lo que implica que el número de representantes del esquema  $H$  (individuos con valor de la aptitud mayor al promedio de toda la población) se incrementa exponencialmente con el tiempo.

### Capítulo 3. Métodos de Optimización con Restricciones.

Un problema de optimización requiere de encontrar un vector  $x \in M$  para el cual, bajo una función  $f: M \rightarrow \mathfrak{R}$  se cumple que  $f(x)$  tiene el valor máximo (o mínimo). Esta función puede representar el costo, el peso o el gusto. Por ejemplo, si la función representa las ganancias y  $M$  es el conjunto de los posibles movimientos a realizar en una bolsa de valores el problema consiste en encontrar un movimiento ( $x$ ) que nos implique las ganancias máximas.

En la mayor parte de los problemas se tiene que no basta con encontrar la solución sobre todos los elementos de  $M$ . Además es necesario cumplir con condiciones adicionales (restricciones) que el propio problema impone sobre la solución. En el ejemplo anterior habrá restricciones como: no es posible invertir más de \$10,000.00 pesos o la inversión mínima es de \$500.00.

Los problemas de optimización han sido atacados, previamente, utilizando programación lineal (cuando la función objetivo y las restricciones son lineales), programación no lineal, programación dinámica, búsqueda tabú, recocido simulado, entre otros.

#### 3.1 Optimización sin restricciones.

Los métodos para resolver problemas de optimización no restringidos se pueden dividir en (a) métodos que utilizan derivadas en la optimización (b) métodos que no usan derivadas y se les conoce como métodos de búsqueda.

Entre los que utilizan derivadas están: Ascenso de máxima pendiente, Newton, Variables conjugadas y otros. A continuación se describen brevemente los dos primeros y mayores detalles se pueden consultar en [7].

##### 3.1.1 Ascenso de máxima pendiente.

El problema general de la programación no lineal sin restricciones se reduce a encontrar:

$$\max[f(x)] \quad x \in M \subseteq \mathfrak{R}^n \quad (15)$$

se tiene que  $\nabla f(x)$  es el vector que apunta en la dirección de mayor crecimiento de  $f$  [22]. De aquí se deriva un procedimiento iterativo para que, dado un vector en el dominio de la función ( $\vec{r}_i$ ), sea posible calcular otro ( $\vec{r}_{i+1}$ ) que también esté en el dominio y  $f(\vec{r}_{i+1}) > f(\vec{r}_i)$ :

$$\vec{r}_{i+1} = \vec{r}_i + \lambda_i \nabla f(\vec{r}_i) \quad (16)$$

donde  $\lambda_i > 0$  y si el valor de  $f(\vec{r}_{i+1}) \leq f(\vec{r}_i)$  se puede reducir el valor de  $\lambda_i$  hasta que se cumpla la condición.

### 3.1.2 Método de Newton.

Este método requiere de la segunda derivada de la función y la forma de transformar  $r_i$  en  $r_{i+1}$  es:

$$r_{i+1} = r_i - [\nabla^2 f(r_i)]^{-1} \nabla f(r_i) \quad (17)$$

donde  $[\nabla^2 f(x_i)]^{-1}$  es el inverso de la matriz Hesiana (la matriz de las segundas derivadas parciales de  $f(x)$  con respecto a  $x_i$ ) evaluadas en  $r_i$ :

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_n^2 \partial x_1} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \quad (18)$$

### 3.1.3 Búsqueda directa.

De los métodos que no utilizan la derivada sólo se expondrá la búsqueda directa; algunos otros se explican en [11]. Este método propuesto por Hooke y Jeeves [14] es una estrategia de búsqueda que utiliza la información ya obtenida y desecha la que no le será útil mediante la topología de la función . En la Figura 6 se muestra un diagrama de este método para minimizar una función  $f$ .

Teniendo el vector  $x$  y la perturbación  $\Delta x$  y a continuación los pasos que se describen:

- Evaluar  $f(x)$  en el punto base, denotado por  $x^{(b)} = x^{(0)}$ .
- Calcular  $x_i^{(1)} = x_i^{(0)} + \Delta x_i^{(0)}$  y evaluar  $f(x + x_i^{(1)})$ . Si el valor de la función en este punto es mayor que el de  $f(x^{(0)})$  éste se adopta como el nuevo vector  $x$ . Si no, se continúa para toda  $i$  hasta que todas las variables independientes han sido cambiadas para completar la búsqueda exploratoria tipo I.
- Después se realiza la búsqueda de patrones a partir del punto  $x$  de acuerdo con la siguiente regla

$$x_i^{(k+1)} = 2x_i^{(k)} - x_i^{(b)} \quad (19)$$

- Con estos nuevos valores se realiza la búsqueda exploratoria tipo II, que consiste en evaluar la función en  $f(x + x_i^{(k+1)})$  para cada  $i$ ; para comparar con los valores anteriormente calculados de la función. Si el mayor de ellos sobrepasa al anterior entonces se cambia el punto base por este nuevo vector, en otro caso se reinicia una búsqueda del tipo I.
- Cuando se han revisado todas estas búsquedas se realiza un decremento sobre  $\Delta x$  como sigue:

$$\Delta x_{i, nueva} = \Delta x_{i, previa} \frac{\Delta x_i^{(0)}}{e^{\epsilon}} \quad (20)$$

donde  $\xi$  es el tamaño del paso con que se realizará la búsqueda

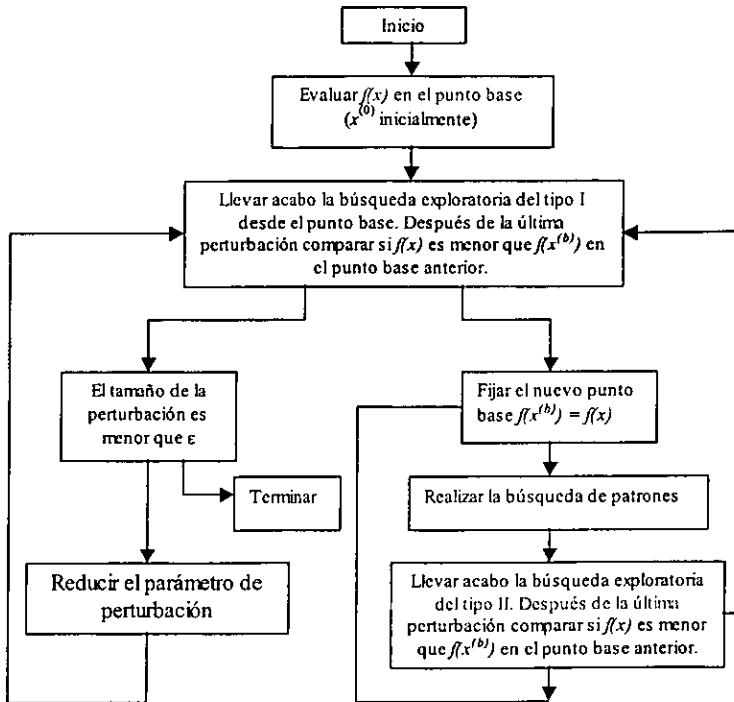


Figura 6. Diagrama de flujo para la minimización con búsqueda directa

Ejemplos de este método se pueden encontrar en [11], así como un análisis comparativo con otros métodos descritos en el mismo libro.

### 3.2 Optimización con restricciones.

Los problemas no lineales con restricciones no lineales son mucho más complicados de abordar que los descritos en la sección anterior. Los métodos propuestos en este caso se puede clasificar como:

- Extensiones de los métodos lineales a los problemas no lineales utilizando aproximaciones.
- Uso de tolerancias flexibles para ajustar tanto vectores factibles como no factibles.

- Transformar el problema restringido y no lineal a un problema sin restricciones, utilizando funciones de penalización.

El material de esta sección ha sido tomado de [11]. El problema general consiste en:

$$\begin{aligned} & \text{Maximizar} && f(x) && x \in \mathfrak{R}^n \\ & \text{Sujeto a} && h_i(x) = 0 && i = 1, \dots, m \\ & && g_i(x) \leq 0 && i = m + 1, \dots, p \end{aligned} \quad (21)$$

### 3.2.1 Aproximación lineal.

Para realizar una aproximación lineal de un problema no lineal es posible hacerlo a través de series de Taylor y considerando sólo el primer orden de la serie. Será necesario repetir este proceso (convertir las funciones a funciones lineales) varias veces para cada problema y se generará una secuencia de soluciones:  $\bar{x}^{(0)}, \bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(k)}$  que, se espera, converge a la solución óptima del problema original.

Entonces el problema (21) se convierte en:

$$\begin{aligned} & \text{Maximizar} && f(\bar{x}^{(k)}) + \nabla^T f(\bar{x}^{(k)})(x - \bar{x}^{(k)}) && x \in \mathfrak{R}^n \\ & \text{Sujeto a} && h_i(\bar{x}^{(k)}) + \nabla^T h_i(\bar{x}^{(k)})(x - \bar{x}^{(k)}) = 0 && i = 1, \dots, m \\ & && g_i(\bar{x}^{(k)}) + \nabla^T g_i(\bar{x}^{(k)})(x - \bar{x}^{(k)}) \geq 0 && i = m + 1, \dots, p \end{aligned} \quad (22)$$

donde  $f(\bar{x}^{(k)})$ ,  $\nabla f(\bar{x}^{(k)})$ ,  $h_i(\bar{x}^{(k)})$ ,  $\nabla h_i(\bar{x}^{(k)})$ ,  $g_i(\bar{x}^{(k)})$  y  $\nabla g_i(\bar{x}^{(k)})$  son vectores constantes o escalares evaluados en  $\bar{x}^{(k)}$ . Por lo tanto este problema se convierte en un problema lineal.

Para que la solución del problema (22) converja al vector solución del problema (21) deben cumplirse las siguientes condiciones:

- Tanto la función como las restricciones deben de ser continuas y diferenciables.
- La función y las restricciones de igualdades debe ser convexas en  $\mathfrak{R}$ .
- Las restricciones con desigualdades deben ser cóncavas.
- La región factible no debe ser vacía, esto es, el problema tiene solución.

Existen muchos métodos para tratar de hacer converger la solución del problema lineal a la solución del problema no lineal. Algunos de ellos se describen en [11].

### 3.2.2 Método de Tolerancia a fallas.

Este método se describe ampliamente en [11]. Consiste en convertir el problema (21) en el siguiente:

$$\begin{aligned} & \text{Maximizar } f(\bar{x}) \quad \bar{x} \in \mathfrak{R}^n \\ & \text{Sujeto a } \Phi^{(k)} - T(\bar{x}) \geq 0 \end{aligned} \quad (23)$$

donde  $\Phi^{(k)}$  es el valor del criterio de tolerancia flexible en el  $k$ -ésimo estado de la búsqueda definida por (21) y  $T(\bar{x})$  es una función positiva de todas las restricciones, tanto igualdades como desigualdades, del mismo problema, usada para medir el grado de violación a las restricciones.

El criterio de tolerancia  $\Phi^{(k)}$  es una función decreciente positiva de los vértices de un poliedro flexible en  $\mathfrak{R}^n$ . Existen varias formas de definirlo, una posible es la siguiente:

$$\begin{aligned} \Phi^{(k)} &= \min \left\{ \Phi^{(k-1)}, \frac{m+1}{r+1} \sum_{i=1}^{r+1} \left| \bar{x}_i^{(k)} - \bar{x}_{r+2}^{(k)} \right| \right\} \\ \Phi^{(0)} &= 2(m+1)t \end{aligned} \quad (24)$$

donde  $t$  es el tamaño del poliedro inicial,  $m$  el número de restricciones de igualdad,  $\bar{x}_i^{(k)}$  es el  $i$ -ésimo vértice del poliedro,  $n$  es la dimensión  $x$ ,  $r = (n - m)$  el número de grados de libertad de  $f(x)$  en el problema (21),  $\bar{x}_{r+2}^{(k)}$  es el centro del poliedro,  $k = 0, 1, 2, \dots$  es un índice de los estados de la búsqueda y  $\Phi^{(k-1)}$  es el valor del criterio de tolerancia en el estado  $(k-1)$  de la búsqueda.

Es posible mostrar que

$$\lim_{x \rightarrow x^*} \Phi^{(k)} = 0 \quad (25)$$

donde  $x^*$  representa la solución del problema (21). La expresión para el criterio de violación de restricciones es el siguiente:

$$T(x) = + \left[ \sum_{i=1}^m h_i^2(x) + \sum_{i=m+1}^p U_i g_i^2(x) \right]^{\frac{1}{2}} \quad (26)$$

donde  $U_i$  es 0 si  $g_i(x) \geq 0$  y 1 para  $g_i(x) < 0$ . Esta función distingue entre los vectores factibles y los no factibles. Si  $T(x^{(k)}) = 0$  entonces  $x^{(k)}$  es factible; en caso contrario no lo será. Para valores pequeños de  $T(x^{(k)})$  la  $x^{(k)}$  estará cerca de ser factible.

De lo anterior y de (25) es claro que el problema (23) resuelve el problema (21). Para una discusión de la manera en que es posible construir los poliedros se puede ver [11].



### 3.2.3 Funciones de penalización.

La esencia de estos métodos es transformar el problema con restricciones no lineal a un problema no restringido o una secuencia de problemas no restringidos. Estos se pueden dividir en métodos paramétricos y sin parámetros.

En el primer grupo se define la siguiente como función de penalización del problema (21):

$$P(\bar{x}^{(k)}, \bar{\rho}^{(k)}) = f(\bar{x}^{(k)}) + \sum_{i=1}^m \rho_i^{(k)} H(h_i(\bar{x}^{(k)})) + \sum_{i=m+1}^p \rho_i^{(k)} G(g_i(\bar{x}^{(k)})) \quad (27)$$

donde  $\rho_i^{(k)} \geq 0$  son los factores de peso,  $H(h_i(\bar{x}^{(k)}))$  y  $G(g_i(\bar{x}^{(k)}))$  son funcionales de  $h_i$  y  $g_i$  y  $k=1,2,\dots$  es el número de estados de la búsqueda.

Tanto los funcionales  $G$ ,  $H$  y los pesos  $\rho$  se pueden escoger libremente pero deben de cumplir con lo siguiente:

- $\lim_{k \rightarrow \infty} \sum_{i=m+1}^p \rho_i^{(k)} G(g_i(\bar{x}^{(k)})) = 0$
- $\lim_{k \rightarrow \infty} \sum_{i=1}^m \rho_i^{(k)} H(h_i(\bar{x}^{(k)})) = 0$
- $\lim_{k \rightarrow \infty} |P(\bar{x}^{(k)}, \bar{\rho}^{(k)}) - f(\bar{x}^{(k)})| = 0$

Esto quiere decir que el efecto de la función de penalización debe ser gradualmente disminuido para que la solución del problema original sea la misma que la de la función de penalización.

Algunos funcionales para  $G$  son:

- $G_1(g_i(\bar{x})) \rightarrow +\infty$  tanto como  $g_i(\bar{x}) \rightarrow 0^+$
- $G_2(g_i(\bar{x})) \rightarrow 0$  tanto como  $g_i(\bar{x}) \rightarrow 0^-$
- $G_3(g_i(\bar{x})) > 0$  para  $g_i(\bar{x}) < 0$  y  $G_3(g_i(\bar{x})) = 0$  para  $g_i(\bar{x}) \geq 0$

en cambio para  $H$  se acostumbra que  $H(h_i(\bar{x})) \rightarrow 0$  si  $h_i(\bar{x}) \rightarrow 0$  y comúnmente se utiliza  $H(h_i(\bar{x})) = h_i^2(\bar{x})$ .

Un caso particular de las funciones de penalización son los multiplicadores de Lagrange. Consiste en resolver el problema (21) transformándolo como se muestra a continuación

$$\begin{aligned} & \text{Maximizar } f(\bar{x}) \quad \bar{x} \in \mathcal{R}^n \\ & \text{Sujeto a } \quad h_i(\bar{x}) = 0 \quad i = 1, \dots, m \\ & \quad \quad \quad g_i(\bar{x}) - v_i^2 = 0 \quad i = m+1, \dots, p \end{aligned} \quad (28)$$

y esto garantiza que las desigualdades del problema se cumplen y todas las restricciones son igualdades.

Entonces se define la función de Lagrange como sigue:

$$P(\bar{x}, \omega) = f(\bar{x}) + \sum_{i=1}^m \omega_i h_i(\bar{x}) + \sum_{i=m+1}^p \omega_i [g_i(\bar{x}) - v_i^2] \quad (29)$$

donde  $\omega_i$  para  $i=1, \dots, p$  son factores de peso independientes de  $\bar{x}$  y se conocen como multiplicadores de Lagrange. Se ha demostrado [3] que es condición necesaria y suficiente para que  $\bar{x}^*$  sea solución de (28) que  $f(\bar{x}^*)$  sea convexa, con un conjunto de restricciones convexo en la vecindad de  $\bar{x}^*$  y que el siguiente conjunto de ecuaciones se satisfaga en  $\bar{x}^*$

$$\begin{aligned} \frac{\partial P(\bar{x}^*)}{\partial x_j} &= 0 & j &= 1, \dots, n \\ \frac{\partial P(\bar{x}^*)}{\partial \omega_i} &= 0 & i &= 1, \dots, p \\ \frac{\partial P(\bar{x}^*)}{\partial v_i} &= 2\omega_i v_i = 0 & i &= m+1, \dots, p \\ \omega_i &\geq 0 & i &= 1, \dots, p \end{aligned} \quad (30)$$

para que el mínimo de  $f(\bar{x})$  ocurra en un punto estacionario de  $P$ .

## Capítulo 4. Métodos Comparados.

Hay muchos métodos para resolver el problema de optimización no lineal restringida utilizando algoritmos genéticos [2],[6] o [24]. Estos se pueden dividir en grupos como sigue:

- Utilizan funciones de penalización
- Mantienen una población de puntos factibles (que cumplen las restricciones) mediante una representación y operadores genéticos especiales.
- Separan el objetivo y las restricciones
- Híbridos
- Con enfoques originales

En esta tesis sólo se analizan algunos métodos de funciones de penalización que son los más difundidos en la literatura. Pero en [2] se describe esta clasificación y varios métodos más.

En los siguientes apartados se describen métodos para resolver el problema de encontrar un vector  $x$  que minimice una función (en los métodos analizados se describe cómo buscar el mínimo, pero también se pueden utilizar para maximizar haciendo un cambio de signo en la función objetivo) y cumpla con:

$$\begin{array}{ll} \text{Minimizar} & f(x) \quad x \in \mathbb{R}^n \\ \text{Sujeto a} & h_i(x) = 0 \quad i = 1, \dots, m \\ & g_i(x) \leq 0 \quad i = m + 1, \dots, p \end{array} \quad (31)$$

este problema se puede transformar en el de minimizar *eval* [24], que se define como sigue:

$$\text{eval}(x) = \begin{cases} f(x) & x \in \text{puntos factibles} \\ f(x) + \text{penalización}(x) & \text{en otro caso} \end{cases} \quad (32)$$

donde la función de penalización es positiva y el conjunto de los puntos factibles se denota con  $M$  y se define como:

$$M = \{x | h_i(x) = 0 \quad i = 1, \dots, m \text{ y } g_i(x) \geq 0 \quad i = m + 1, \dots, p\} \quad (33)$$

Se busca construir la función de penalización tal que la solución de (32) sea la solución de (31). Si  $\bar{x}_j$  es la solución (31) entonces se cumple que:

$$f(\bar{x}_j) \leq f(\bar{x}) + P(\bar{x}) \quad (34)$$

donde  $P(x)$  es la función de penalización.

En el resto de la tesis se utiliza para definir la función de penalización a  $f_j$  ( $1 \leq j \leq m$ ) que mide la violación a la  $j$ -ésima restricción (como se hace en [25]).

Esto es  $f_j(\bar{x})$ , es cero si  $\bar{x}$  satisface la restricción y mayor que cero si  $\bar{x}$  no la satisface. Se define como sigue:

$$f_j(\bar{x}) = \begin{cases} \max\{0, g_j(\bar{x})\} & 1 \leq j \leq m \\ |h_j(\bar{x})| & m+1 \leq j \leq p \end{cases} \quad (35)$$

Los distintos métodos que se describen a continuación hacen uso de esta definición para construir la función de penalización.

#### 4.1 Método de Homaifar et al.

Este método se ha descrito en [13] y la función de penalización que se propone es la siguiente:

$$P(\bar{x}) = \begin{cases} 0 & \bar{x} \in M \\ \sum_{i=1}^m R_{i,j} g_i^2(\bar{x}) & \bar{x} \notin M \end{cases} \quad (36)$$

La constante  $R$  tomará distintos valores dependiendo de  $g_i$  y el número  $l =$  niveles de violación definidos. Por ejemplo si,  $a \leq g_i(\bar{x}) \leq b$  el intervalo  $a, b$  será dividido en  $l$  partes y así se obtienen un conjunto de  $a_j$  con  $j = 0, 1, \dots, l$  que son extremos de cada división (donde  $a_0 = a$  y  $a_l = b$ ) y dependiendo de en qué subintervalo esté el valor de  $g_i(\bar{x})$  se tomará la  $R_{ij}$  correspondiente.

La penalización que se describe en [13] sólo considera a  $f_i$  como  $g_i^2(\bar{x})$  y no como la define en (35). Para hacer la implementación y poder comparar con el resto de los métodos se tomó la siguiente definición de  $P(\bar{x})$ , que es equivalente a (36) en el caso de no haber restricciones de igualdad.

$$P(\bar{x}) = \begin{cases} 0 & \bar{x} \in M \\ \sum_{i=1}^p R_{i,j} f_i^2(\bar{x}) & \bar{x} \notin M \end{cases} \quad (37)$$

La constante  $R$  que aparece en la definición tiene dos índices porque uno refiere al nivel de la violación que se aplicará ( $j$ ) y el otro la restricción violada ( $i$ ). Entonces es necesario definir tanto los valores de las constantes  $R$  como los extremos de cada rango de violación ( $a_j$ ) y el número de niveles (rangos) de violación que se tendrán ( $l$ ).

$$R_{ij} = \begin{cases} R_{i,1} & \text{si } a_{0,i} < f_i(\bar{x}) < a_{1,i} \\ R_{i,2} & \text{si } a_{1,i} < f_i(\bar{x}) < a_{2,i} \\ \dots & \dots \\ R_{i,l} & \text{si } a_{l-1,i} < f_i(\bar{x}) < a_{l,i} \end{cases} \quad (38)$$

Así, la  $R$  utilizada depende del valor de la violación. Es necesario definir  $ml$  valores de  $R_{ij}$  más  $m(l+1)$  valores de  $a_j$ ; en total  $m(2l+1)$  valores;

Homaifar et al. afirman que no es tarea fácil la selección de los valores de  $R$  [13] y que el desempeño del AG dependerá de la selección de éstos. En su artículo muestra tres ejemplos con distintos valores de  $R$  y  $a$ . Pero en el primero (que se muestra en el apéndice I como función 7) el método produce un error que se describe a continuación.

Los parámetros que utiliza son  $l = 5$  (cinco niveles de penalización),

$$R_{ij} = \begin{vmatrix} 500 & 0 \\ 100 & 0 \\ 50 & 150 \\ 100 & 190 \\ 500 & 500 \end{vmatrix} \text{ y } a_{ij} = \begin{vmatrix} -\infty & -\infty \\ -0.1 & -1 \\ -0.01 & 0 \\ 0.01 & 0.01 \\ 0.1 & 0.1 \\ \infty & \infty \end{vmatrix} \quad (39)$$

La solución que encuentra es  $x_h = (0.832031, 0.910156)$  y  $f(x_h) = 1.4339$ . Este punto no cumple con la primera restricción (tiene un error de 0.03712). Pero al momento de hacer la implementación de este método, utilizando los parámetros que describe el autor, el AG es capaz de encontrar un vector  $x_h = (0.8080, 0.88544)$  con  $f(x_h) = 1.387011995830$ . Este punto tampoco satisface la primera restricción, tiene un error de 0.01, pero el valor que encuentra de la función es menor y ¡menor aún que la solución exacta del problema! La solución exacta al problema es

$$x_s = \left( \frac{-4 + \sqrt{112}}{8}, \frac{4 + \sqrt{112}}{16} \right) \approx (0.8228756555, 0.9114378277) \quad (40)$$

$$f(x_s) = \frac{288 - 23\sqrt{112}}{32} \approx 1.3934649806$$

Esta función de penalización, con los valores de  $R$  y  $a$  descritos, no cumple con la ecuación (34). Esto es, la función de penalización no es lo "suficientemente buena" como para castigar a un punto que no cumple con las restricciones y es menor que la solución.

La suma del valor de la función más la función de penalización se muestra en la Tabla 4.

Tabla 4. Valores que llevan a un valor equivocado en el ejemplo 1 de [13]

Elemento	Valor
$g_1 = 0$	0.011719
$g_1 > 0$	-0.001453
$R_1 g_1(x_h) \quad R_1 = 100$	$100 * (0.011719)^2 = 0.0137334961$
$R_2 g_2(x_h) \quad R_2 = 500$	$500 * (-0.001453)^2 = 0.0010556045$
$f(x_h)$	1.372222900391
$F(x_h) = f(x_h) + P(x_h)$	1.387011995830

Este error se puede corregir haciendo los números  $R_{ij}$  más grandes para que al multiplicar por el cuadrado de la restricción no sea pequeño el castigo. En el ejemplo anterior este error se corrige si se utilizan constantes 10 veces mayores. Así se puede obtener un error de 0.0061 y el vector cumple ambas restricciones.

En [24] se afirma que si los coeficientes  $R$  son moderados, el algoritmo puede converger a una solución no factible y si son demasiado grandes el método es equivalente a rechazar las soluciones no factibles.

Este error está causado porque el castigo a los puntos no factibles puede ser muy pequeño cuando se está cerca de la región de factibilidad (problema que presentan todas las penalizaciones estáticas). Esto se ilustra en la Figura 7 donde se tiene una función  $f: \mathcal{R} \rightarrow \mathcal{R}$  con la restricción  $a < x < b$ . La solución al problema está en un punto dentro del intervalo ( $x_s$ ), pero el valor de la función fuera del intervalo (en la región no factible) más la función de penalización es menor que la solución del problema.

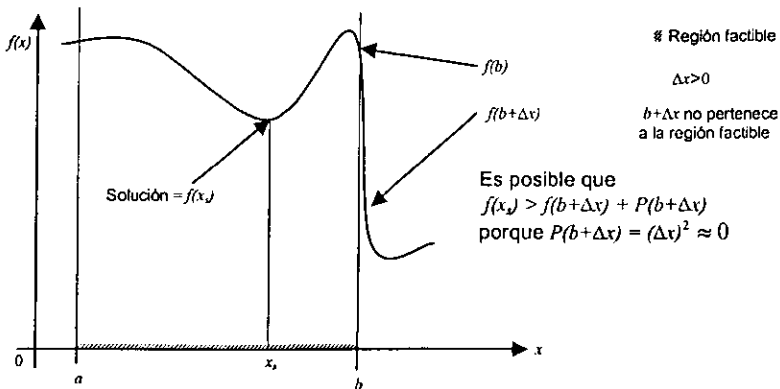


Figura 7. Ejemplifica la causa de un posible error en la función de penalización

El error descrito en la Tabla 4 no es exactamente como el mostrado en la figura anterior. En el caso de la función 7 del apéndice se puede ver que no tiene esta clase de "caídas" exponenciales (se trata de un paraboloide). El problema consiste en que la restricción de igualdad describe una recta en el plano y la desigualdad una elipse. Los puntos factibles están sobre la recta y dentro de la elipse. La solución al problema es uno de los puntos de intersección entre la elipse y la recta. El error está causado porque un punto cercano siempre puede alejarse de la solución para obtener un valor menor. Esto se muestra en la Figura 8 donde se han dibujado dos rectas. La línea continua en el plano  $x_1 \times x_2$  representa a la restricción de igualdad y la punteada a otra recta que tiene un valor de la función de objetivo menor. El punto solución es el vector  $\bar{x}_s$  (representado por la línea —), pero el AG encontrará un punto cuya penalización sea pequeña (por estar

cerca de cumplir con la restricción, es decir, cerca de la línea continua en el plano  $x_1 \times x_2$ ) y el valor de la función objetivo sea menor.

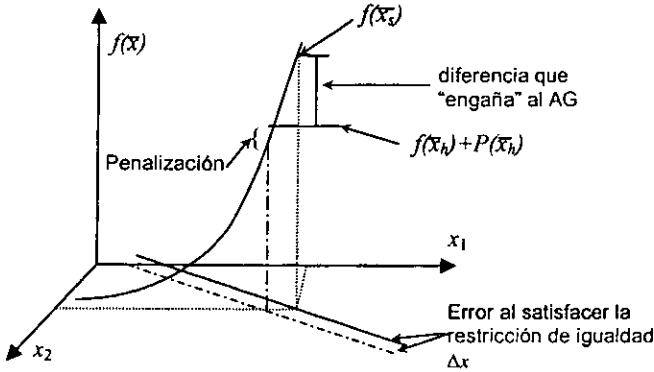


Figura 8. Causa de un posible error en la función de penalización

En los otros dos ejemplos reportados en [13] no se presenta este problema, pero sí hay algunos de copia de las funciones, por lo cual fue necesario consultar la fuente original del problema [11] (esto también lo reportaron en [18]).

En el análisis comparativo que se realizó en este trabajo no fue posible tomar valores de  $R_{ij}$  especiales para cada función por lo que se decidió utilizar cuatro niveles de penalización con valores de  $R = 100, 200, 500, 1000$  e intervalos de  $(0, 10)$ ,  $(10, 100)$ ,  $(100, 1000)$  y  $(1000, \infty)$ . Estos valores se encontraron después de hacer pruebas con las funciones del apéndice I. Estas pruebas consistieron en dejar los otros parámetros fijos y variar sólo los valores de  $R$  y  $l$  buscando con cuál de ellos se obtenía un menor valor para la aptitud en un número fijo de generaciones.

#### 4.2 Método de Joines y Houck.

Este método se comporta como una función de penalización no estacionaria (o dinámica) y se describe en [17]. Se define como sigue:

$$P(\bar{x}, \alpha, \beta) = \rho_k^\alpha \times SVC(\beta, \bar{x})$$

$$\rho_k = C \times k \quad k = \# \text{generación} \quad (41)$$

$$SVC(\beta, \bar{x}) = \sum_{i=1}^p f_i^\beta(\bar{x}) \quad \beta = 1, 2, \dots$$

donde  $\alpha, \beta$  y  $C$  son parámetros del método y  $k$  es el número de la generación que se está calculando. Define, también, otra función de penalización ( $R$ ) como sigue:

$$R(\alpha, \beta) = e^{P(\alpha, \beta)} \quad (42)$$

Los valores que se utilizaron para los parámetros son  $C = 0.5$ ,  $\alpha = 2$  y  $\beta = 2$ . No se utilizó la función  $R$  porque ésta causaba desbordamiento en las variables decimales que se utilizaron (la función alcanza números del orden de  $10^{600}$  muy rápidamente y el límite del software utilizado es  $10^{400}$ ).

Con esta función de penalización no es posible caer en el problema del método de Homaifar porque si el valor de  $SVC$  (por sus iniciales en inglés *sum of the violated constraints*) se hace muy pequeño en las siguientes generaciones será castigado con mayor rigor.

Esta característica hace que aún cuando se tenga un AG elitista el mejor individuo de la generación actual no sea igual de apto en la siguiente. Una gráfica del valor de la aptitud en cada generación (ver Apéndice III) puede comportarse como se muestra a continuación:

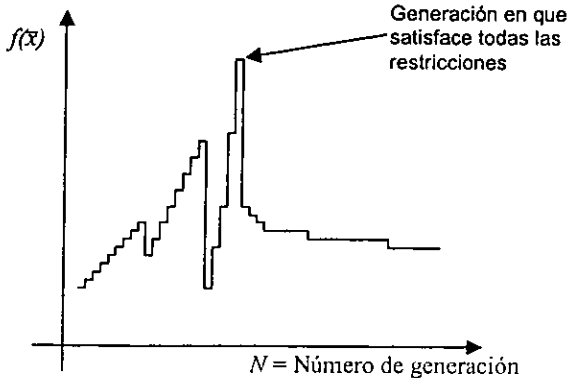


Figura 9. Gráfica del valor de la aptitud del mejor individuo contra número de generación.

En la Figura 9 se puede apreciar que aún cuando se trata de minimizar la función el valor del mejor individuo de una generación a la otra: ¡crece! Esto es provocado por un individuo que no satisface todas las restricciones (el mejor en la generación  $i$ ) que es castigado más fuertemente en la siguiente generación y su aptitud será mayor, pero aún así se conserva como el mejor individuo. La gráfica de esta función siempre da "brincos" (crece y decrece en alguna generación, luego sigue creciendo y decreciendo) hasta que encuentra una posible solución que satisface todas las restricciones. De este momento en adelante se convierte en una función decreciente (como cualquier AG elitista).

La implementación realizada en este trabajo no reproduce los resultados del autor (esto debido a que el utiliza una selección de rango geométrica, cruce aritmética y otros operadores que no se trabajaron en esta tesis), pero si encuentra soluciones semejantes. En la Tabla 5 se muestra un ejemplo para el primer ejercicio reportado por Joines.



Tabla 5. Valores reportados en [17] y calculados en este trabajo.

Encontrado con $P(2,2)$	Reportados para $P(1,1);P(1,2);P(2,2);R(1,1)$
-6961.474826	-6961.7853;-7028.0926;-6961.2399;-6961.8253

En los resultados reportados en [17] sólo el  $P(1,1)$  y  $P(2,2)$  satisfacen las restricciones, los otros no. El calculado en este trabajo también cumple ambas restricciones.

#### 4.3 Método de Schoenauer y Xanthakis.

Este métodos ha sido llamado "*behavioural memory*" y fue propuesto en 1993 [29]. Se basa en la idea de evaluar las restricciones en un orden en particular.

Los primeros pasos del algoritmo tratan de definir la región factible y, cuando ya se tiene ésta, entonces se optimiza la función objetivo.

El algoritmo empleado es el siguiente [29] [25],[24]:

- Iniciar con una población aleatoria (con individuos factibles y no factibles)
- Fijar  $j = 1$  ( $j$  es contador de restricciones)
- Evolucionar esta población para minimizar la violación a la restricción  $j$ -ésima hasta que un porcentaje de la población ( $\phi$ ) es factible para esta restricción. En este caso se tiene:

$$eval(\vec{x}) = g_j(\vec{x}) \quad (43)$$

- Incrementar  $j$
- La población actual es el punto de partida de la siguiente fase de la evolución, que consiste en minimizar la violación de la  $j$ -ésima restricción. Durante esta fase los puntos que no satisfacen las restricciones  $1, 2, \dots, j$ -ésima son eliminados de la población. El criterio para detener este procedimiento consiste en tener un porcentaje de la población ( $\phi$ ) que satisface la restricción  $j$ -ésima.
- Si  $j < m$ , se repiten los dos pasos anteriores, en otro caso ( $j = m$ ) se optimiza la función  $f$  rechazando todos los individuos no factibles.

Para la implementación de este método fue necesario que se utilizara elitismo total, porque de otra forma el número de individuos en la población se disminuía considerablemente cada vez que se obtenía una nueva generación. Al conservar a los mejores  $N$  individuos ( $N$  tamaño de la población) se garantiza que aún cuando no se obtenga ningún nuevo individuo que cumpla la restricción  $j$  siempre se cuenta con todos los individuos de la generación anterior. Esto hace que el tamaño de la población permanezca constante al eliminar a los individuos que violan las restricciones.

El orden en que se evalúan las restricciones cambia el resultado de la función [25]. Las restricciones de igualdad ( $h_i$ ) se evalúan al final, porque es difícil que el AG encuentre combinaciones de las variables que cumplan con éstas.

#### 4.4 Método de Powell y Skolnick.

Este método fue desarrollado alrededor de 1993 [27] e incorpora una regla heurística para distinguir los puntos no factibles: "Toda solución factible es mejor que una solución no factible" [24]. La función de penalización se define como sigue:

$$P(\bar{x}) = r \sum_{j=1}^p f_j(\bar{x}) + \rho(\bar{x}, t) \quad (44)$$

$$\rho(\bar{x}, t) = \max \left\{ 0, \max_{x \in M} \{f(\bar{x})\} - \min_{x \in M} \{f(\bar{x}) + \sum_{j=1}^p f_j(\bar{x})\} \right\}$$

Esto es, los individuos no factibles incrementan la penalización, porque no pueden ser mejores que el peor de los individuos factibles.

Al hacer la implementación se toman todos los individuos de la población y se construye el conjunto de los factibles y después el de los no factibles. De aquí se calcula  $\rho$ , mientras no hay puntos factibles en la población el castigo está dado sólo por el primer factor, pues el conjunto de los no factibles hace que  $\rho$  sea cero. Al aparecer puntos factibles esto cambia y se castiga más a los individuos que violan las restricciones. En las pruebas realizadas en este trabajo se usó un valor de  $r = 2.$ , valor que se obtuvo de realizar algunos experimentos con las funciones descritas en Apéndice I.

#### 4.5 Método de Kuri.

Este es uno de los métodos más sencillos y en [2] se le ha llamado "Death penalty" porque no aparece explícitamente en la función de penalización la magnitud con la que se ha violado una restricción (la función  $f_i$  definida en (35)) sino solamente la cantidad de restricciones violadas. No es pena de muerte en el sentido de que rechaza a los individuos no factibles como se menciona en [25].

La penalización está definida como sigue:

$$P(x) = K - \sum_{i=1}^s \frac{K}{p} \quad (45)$$

donde  $K$  es una constante grande (del orden de  $10^9$ ),  $s$  es el número de restricciones cumplidas y  $p$  el total de restricciones.

Este método tiene problemas cuando no se tiene individuos factibles en la población, porque todos son penalizados de la misma forma y no se proporciona suficiente información al AG para buscar individuos factibles. Esto se ha resuelto

iniciando el AG con una población de individuos factibles que se busca utilizando el mismo AG [21].

El problema principal para este método son las restricciones con igualdades. Estas se pueden abordar haciendo un cambio en la función objetivo como sigue:

$$\begin{array}{l} \min \{f(\bar{x})\} \\ \text{sujeto } h_i(\bar{x}) = 0 \end{array} \Rightarrow \begin{array}{l} \min \{f(\bar{x}) + h_i(\bar{x})\} \\ \text{sujeto } h_i(\bar{x}) \geq 0 \end{array} \quad (46)$$

donde  $i$  es el índice de las restricciones con igualdad y esto permitirá al algoritmo encontrar valores factibles más fácilmente (éste es una afirmación intuitiva y será necesario hacer un análisis estadístico para comprobar su veracidad). En la implementación de este algoritmo no se consideró esta forma de tratar las restricciones de identidad ni la inicialización de la población con individuos factibles.

## Capítulo 5. Análisis comparativo.

Para poder realizar el análisis de los métodos descritos en el Capítulo 4 se apeló al teorema del límite central y a la prueba de bondad de ajuste de  $\chi^2$ . En este capítulo se hace una descripción de estas técnicas estadísticas sin entrar en gran detalle. En las referencias proporcionadas se pueden consultar las demostraciones y ejemplos. Posteriormente se describen los experimentos desarrollados y cómo se utilizaron las herramientas estadísticas para realizar la comparación entre los distintos métodos.

### 5.1 Sustento estadístico.

Cuando se desea determinar, en forma estadística, el comportamiento de alguna de las características de una población se realiza un muestreo de los individuos que pertenecen a ésta y se construye una tabla de distribución de probabilidades. Esto es, se encuentra la probabilidad de obtener el valor de una característica como resultado de una medición sobre la población. A la función que asocia a cada valor posible con su probabilidad se le llama función de distribución de probabilidades. Entre las funciones de distribución más conocidas están la normal,  $\chi^2$ , uniforme, Poisson, [15]. A continuación se describen las dos primeras.

#### 5.1.1 Distribución Normal

Una distribución común es la distribución normal. Su función de densidad de probabilidad (fdp) se define con dos parámetros  $\mu$  y  $\sigma$  por medio de la siguiente relación:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty, \sigma > 0 \quad (47)$$

La distribución normal es una curva con forma de campana simétrica alrededor  $\mu$  y los valores de  $\mu$  y  $\sigma$  corresponden al promedio y la varianza respectivamente (ver Figura 10). Una de las características interesantes de esta distribución es que 2/3 de las observaciones están en el intervalo  $\mu \pm \sigma$ .

Para obtener la probabilidad asociada con un intervalo específico  $(a, b)$ , de una variable aleatoria con distribución normal, se debe de calcular lo siguiente:

$$P(a < x < b) = \int_a^b f(x) dx \quad (48)$$

Para hacer el cálculo de estas probabilidades se utiliza la simetría de la distribución y se calcula la función de distribución de probabilidades (FDP):

$$F(x) = \int_{-\infty}^x \frac{e^{-\frac{(t-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dt \quad (49)$$

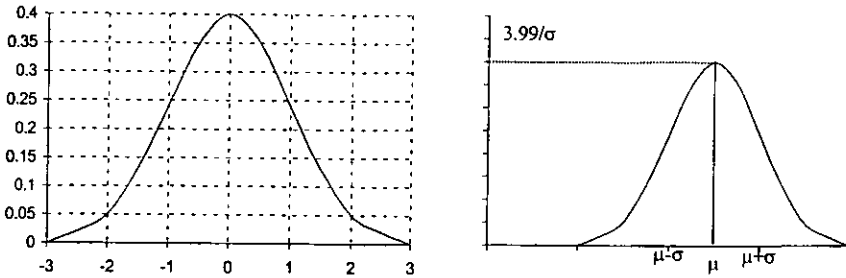


Figura 10. fdp normal con  $\mu = 0$  y  $\sigma = 1$  (izquierda) y con valores arbitrarios (derecha)

Cuando  $\mu = 0$  y  $\sigma = 1$  se le llama distribución normal estándar. Esta distribución ha sido tabulada para facilitar los cálculos de la probabilidad y si se tiene una media  $\mu$  y una desviación  $\sigma$  se puede realizar un cambio de variable para obtener la distribución normal estándar.

$$z = (x - \mu) / \sigma \quad (50)$$

Los parámetros de la distribución nos dan información sobre dónde está el centro de los datos ( $\mu$ ) y cuánto se dispersa ( $\sigma$ ). Al realizar un muestreo aleatorio de una población que tiene una distribución normal se pueden estimar  $\mu$  y  $\sigma$  a partir de  $\bar{X}$  y de  $s$  que se definen a continuación:

$$\bar{X} = \frac{\sum x_i}{n} \quad y \quad s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} \quad (51)$$

a  $s$  se le conoce como el error estándar de la muestra y a  $\bar{X}$  como el promedio.

### 5.1.2 Distribución $\chi^2$

La distribución de  $\chi^2$  es generalmente usada para realizar pruebas de hipótesis en intervalos de confianza predefinidos. Esta distribución depende del parámetro  $\nu$  conocido como "grados de libertad" y su fdp esta dado por:

$$f(x) = \frac{e^{-x/2} \cdot x^{\nu/2 - 1}}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \quad x > 0 \quad (52)$$

donde  $\nu$  es un entero positivo, y  $\Gamma$  es la función gama. En las tablas donde se calculan los valores de esta distribución se utiliza el término de valor crítico. Éste refiere al valor que debe tomar la variable para que el área bajo la curva y a la derecha sea igual a  $\alpha$ . Se les llaman críticos porque determinan la frontera para que se acepte o se rechace una hipótesis [16].

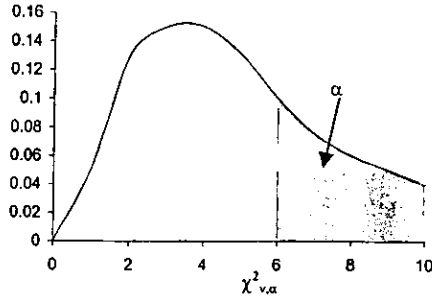


Figura 11. Gráfica de  $\chi^2$  con sus parámetros (grados de libertad y área bajo la curva)

Una gráfica de esta distribución se muestra en la Figura 11 los valores del área debajo de la curva se muestra para 7 grados de libertad en la Tabla 6.

Tabla 6. Valores críticos de la distribución  $\chi^2$

0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
0.990	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3

### 5.1.3 Teorema del Límite Central.

Para poder caracterizar a una población es posible usar una distribución muestral. Esto es, se toman varias muestras de tamaño  $n$  de una misma población y se calcula para cada una de ellas algún estadístico muestral (promedio o error estándar). El conjunto de estos estadísticos forman la distribución muestral. El valor medio de la distribución muestral de promedios se designa con  $\mu_x$  y el error estándar correspondiente con  $\sigma_x$ .

El teorema del límite central afirma lo siguiente:

“Si se toman muestras aleatorias de tamaño  $n$  de una población con media  $\mu$  y desviación estándar  $\sigma$  la distribución muestral de promedios de las  $X$  será una distribución aproximadamente normal. La media de la distribución de muestreo ( $\mu_x$ ) será igual a  $\mu$  y la desviación estándar de la distribución de muestreo ( $\sigma_x$ ) será igual  $\sigma/\sqrt{n}$  (A veces  $\sigma_x$  se llama error estándar de la media). La aproximación será mejor al tomar muestras de mayor tamaño”. [16].

Es importante notar que el teorema anterior no requiere que la distribución original sea normal.

El teorema anterior afirma que la distribución muestral tiende a ser normal, pero es necesario tener una prueba cuantitativa para determinar, con una probabilidad dada, cuándo una distribución es normal o no. A este tipo de pruebas se les

conoce como de bondad de ajuste o de contraste (*goodness-of-fit test*). Se utilizan comúnmente para determinar si es aceptable suponer el ajuste a una cierta distribución. Algunas de las pruebas más conocidas son: K-S (*Kolmogorov-Smirnov*), gráfica, Q-Q Plot (*quantile-quantile plot*) [15], *Anderson-Darling* [26] y  $\chi^2$  [8]. A continuación se describe la última de éstas.

#### 5.1.4 Prueba de $\chi^2$

La prueba de ajuste de  $\chi^2$  se puede aplicar para verificar qué distribución tienen los datos (Poisson, Normal, uniforme, etc.). Dependiendo del número de parámetros que tenga la distribución se tomará un grado de libertad más. Aquí sólo se describe el caso de la distribución normal, es decir,  $\nu = p + 1 = 2 + 1$  (porque la distribución normal depende de dos parámetros)

Supóngase que se tiene una muestra de  $N$  datos para la cual se desea saber si es aceptable o no asumir que tienen una distribución normal. Para resolver este problema se proponen las siguientes hipótesis:

$H_0$ : La función de densidad está dada por (47). La distribución es normal.

$H_1$ : La función de densidad no está dada por (47). La distribución no es normal.

Para decidir cuál de ellas es verdadera se divide el dominio de la variable aleatoria normal en  $k$  intervalos excluyentes,  $(-\infty, a_1)$ ,  $(a_1, a_2)$ , ...,  $(a_{k-1}, +\infty)$  y se calcula la probabilidad de que una variable aleatoria normal esté dentro de estos intervalos:  $p_1, p_2, \dots, p_k$ . Para calcular estas probabilidades se utiliza  $\bar{X}$  y  $s$  (de los  $N$  datos) en lugar de  $\mu$  y  $\sigma$  que generalmente son desconocidas. A continuación se calcula lo siguiente:

$$\chi_{k-3}^2 = \sum_{i=1}^k \frac{(Np_i - x_i)^2}{Np_i} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (53)$$

donde  $x_i$  es el número de datos que pertenecen al intervalo  $i$  (frecuencia observada  $O_i$ ),  $Np_i$  es el número de datos esperados ( $E_i$ ) en el mismo intervalo en una distribución normal.

El estadístico definido en (53) se distribuye aproximadamente como  $\chi_{k-3}^2$  si la hipótesis nula es verdadera. Entonces la hipótesis nula ( $H_0$ ) se rechaza si  $\chi_{k-3}^2 > \chi_{k-3, 1-\alpha}^2$ , donde  $1-\alpha$  determina el intervalo de confianza de la prueba. Normalmente se utiliza  $\alpha=0.1$  o  $0.05$  para un intervalo de confianza del 90 y 95% respectivamente.

Si la hipótesis nula no se rechaza entonces la probabilidad de que datos analizados no tengan una distribución normal es de 10% (o 5% dependiendo del valor de  $\alpha$  seleccionado).

El uso de esta prueba de bondad de ajuste no se recomienda cuando alguno de los intervalos  $(a_i, a_{i+1})$  posea un valor observado pequeño. Lo recomendable es que se tengan 5 datos en cada intervalo como mínimo. Para el desarrollo de las pruebas presentadas a continuación se tomó este último criterio. Se tomaron intervalos con una probabilidad (área bajo la curva de distribución normal) de 0.1 y se comenzó con 50 muestras de tamaño 36 (esto para tener 10 intervalos y cada uno con 5 datos como mínimo). Por lo anterior no se realiza el cálculo del estadístico (53) si existe  $O_i < 5$  para alguna  $0 < i < 11$  y se requiere hacer más muestras (de tamaño  $n=36$ ) hasta que se satisfaga esta condición.

### *5.2 Diseño y desarrollo de los experimentos.*

El propósito de esta tesis es comparar los métodos descritos en el Capítulo 4. Para ello se han realizado distintos experimentos y las pruebas estadísticas descritas en la sección anterior.

El análisis estará centrado en encontrar el comportamiento promedio y desviación estándar de cada método para algunas funciones (descritas en el Apéndice I). Cabe mencionar que no ha sido posible hacer el análisis para cualquier función, porque para comparar cuál método ha encontrado un mejor valor es necesario conocer la solución del problema y al realizar una selección aleatoria de problemas de optimización restringida no se tendría su solución.

El procedimiento consistió en tomar uno de los métodos y probar su comportamiento con alguna de las 25 funciones (se hace una selección al azar) y con una semilla dada del generador de números pseudo-aleatorios.

El procedimiento anterior se ha realizado 2340 ( $= 65 \times 36$ ) veces y se crean 65 ( $N$ ) grupos de 36 ( $n=36$ ). Para cada grupo de 36 se calcula el promedio y se tiene una distribución de promedios (de  $N$  datos). A esta nueva población se le calcula su promedio y error estándar ( $\mu_x, \sigma_x$ ) para realizar la prueba de  $\chi^2$  y determinar si se tiene una distribución normal o no. En caso afirmativo se tiene caracterizado al método; de lo contrario se calculan otros 180 ( $5 \times 36$ ) experimentos y se agregan a la muestra para determinar si cumplen la prueba de bondad de ajuste o si es necesario agregar otros 180.

Antes de hacer el promedio de los datos se han normalizado las funciones dividiendo el valor encontrado entre la solución del problema. Esto es necesario para poder comparar entre distintas funciones. Si el valor óptimo de una función es 100 y el de otra es 0.5 y el método utilizado alcanza los valores óptimos entonces el promedio es 50.25 y este resultado no proporciona información sobre el comportamiento del método. Pero si ambos se dividen entre la solución el promedio será 1. Este método de normalización de las funciones facilita detectar cuántas de las funciones analizadas alcanzaron un valor óptimo: entre más 1's existan será mejor el método.



Como algunas de las funciones que se han planteado para prueba son bastante complicadas, en 100 generaciones ninguno de los métodos obtendrán un resultado que cumpla todas las restricciones. Se ha tomado como parte de la normalización que las soluciones que no cumplan con todas las restricciones del problema se les asigne el siguiente valor:

$$Max \times (1 - s / p) \quad (54)$$

donde *Max* es el mayor valor de la aptitud encontrado en la última generación en todas las pruebas.

Esta asignación impedirá que se tengan números menores a 1 como resultado de algún experimento. Porque si alguno de los métodos presentan el problema descrito para el método de Homaifar (sección 4.1), donde se reporta un valor menor a la solución del problema, se le asignará el valor de (54). También se evita tener números negativos (en el caso de que un método encuentre un valor negativo que no cumpla las restricciones, al dividir por la solución se tendría un valor negativo).

Como se describió en el Capítulo 2 el AG tiene algunos parámetros que pueden ser cambiados y estos afectan su desempeño. Es posible que algunos de los métodos comparados tengan un mejor comportamiento para un conjunto de parámetros y peor para otros. Para poder determinar si esto sucede se han realizado los experimentos que se describen en la Tabla 7.

Tabla 7. Valores de los parámetros del AG que se utilizarán

Paramento de AG	Experimentos			
	G1	G2	G3	G4
Dígitos decimales binarios	20	20	20	20
Dígitos enteros binarios	14	14	14	14
Método de codificación	Gray	Gray	Gray	Gray
Probabilidad de Mutación	0.09	0.05	0.07	0.07
Probabilidad de Cruza	0.25	0.90	0.90	0.90
Tamaño de la población	100	100	50	50
Tipo de selección	Ruleta	Ruleta	Ruleta	Vasconcelos
Tipo de Cruza	Punto Fijo	Punto Fijo	Punto Fijo	Uniforme
Elitismo	1	1	25	Total
Número de generaciones	100	100	100	100

Los parámetros que se presentan no tienen la finalidad de ser exhaustivos. Por el contrario, solamente son una muestra para verificar si el comportamiento de algún método cambia con distintos parámetros. Para asegurar que la eficiencia de los métodos analizados no depende de los parámetros del AG o mostrar cuál es la dependencia es necesario realizar un muestreo estadístico que abarque el espacio de todos los posibles valores de los parámetros y esto requiere de un trabajo distinto.

### 5.3 Resultados.

En la Tabla 8 se muestra el número de experimentos que se realizaron para cada una de los distintos parámetros empleados. Sólo en los parámetros G4 aparece el

método de Schoenauer, esto es, porque la implementación que se realizó de este requiere de elitismo total y en el resto de los "G's" no se utilizó este tipo de elitismo.

Tabla 8. Número de experimentos realizados y valor de la prueba  $\chi^2$ 

Parámetros usados (Tabla 7)	Método	Experimentos	Valor de $\chi^2$
G1	Kuri	90*36=3240	6.22222
	Homaifar	85*36=3060	1.00000
	Joines	90*36=3240	2.44444
	Powell	97*36=3492	9.08247
G2	Kuri	80*36=2880	10.00000
	Homaifar	110*36=3960	8.54545
	Joines	95*36=3420	3.84848
	Powell	110*36=3960	12.36363
G3	Kuri	75*36=2700	4.33333
	Homaifar	120*36=4320	10.33333
	Joines	80*36=2880	5.00000
	Powell	100*36=3600	7.55555
G4	Kuri	86*36=3096	8.18181
	Homaifar	100*36=3600	4.66666
	Joines	100*36=3600	7.88888
	Powell	75*36=2700	3.00000
	Schoenauer	125*36=4500	3.24242

Algunos métodos se repitieron un número de veces que no corresponde a un múltiplo de 180 =5\*36 (en especial G2-Powell que se realizó con 97 y G4-Kuri con 86) debido a que el programa implementado se detuvo antes por razones ajenas al experimento (se congeló la máquina o fallos de energía eléctrica). Pero al calcular la prueba con los datos obtenidos pasaron la prueba estadística y no se agregaron más datos.

En total se ejecutaron 1,618\*36=58,248 experimentos. Cada uno de ellos tardaba 2.5 minutos en promedio lo que da un total de 15 semanas de cómputo sin interrupción.

Los valores de  $\mu_x, \sigma_x$  se muestran en la Tabla 9. Tanto el promedio muestral como el error estándar de las medias ( $\mu_x, \sigma_x$ ) tienen un orden de 1E280 porque la constante utilizada para evitar el error descrito en la sección 4.1 tiene un valor de 5.0E281. Este es el máximo valor en la última generación de todos los experimentos con todos los parámetros utilizados. Es un valor muy grande porque el método de Powell (que arroja este valor) siempre calcula el máximo error para castigar a los individuos que no cumplen las restricciones.

Tabla 9. Valores de  $\mu_x, \sigma_x$  para los experimentos realizados

Parámetros usados (Tabla 7)	Método	$\mu_x$	$\sigma_x$
G1	Kuri	1.65991E281	2.30415E280
	Homaifar	2.35829E281	2.76944E280
	Joines	2.33657E281	2.70489E280
	Powell	2.13516E281	2.53343E280
G2	Kuri	1.47155E281	2.20060E280

Parámetros usados (Tabla 7)	Método	$\mu_x$	$\sigma_x$
	Homaifar	2.37554E281	2.58420E280
	Joines	2.32140E281	2.35599E280
	Powell	2.13868E281	2.22789E280
G3	Kuri	1.66157E281	2.05705E280
	Homaifar	2.40939E281	2.14969E280
	Joines	2.29502E281	2.41732E280
	Powell	2.07283E281	2.34106E280
G4	Kuri	1.61322E281	2.00831E280
	Homaifar	2.27267E281	3.54317E280
	Joines	2.14620E281	2.59664E280
	Powell	1.90579E281	2.39071E280
	Shoenauer	1.92934E281	2.40484E280

#### 5.4 Comentarios a los resultados

Con los datos de la Tabla 9 es posible calcular tanto el peor de los comportamientos que tendrá el método como el mejor de ellos y ordenar de acuerdo a estos datos para encontrar aquel que tiene un mejor desempeño y compararlo contra el que tenga el peor desempeño. Estos se muestran en la Tabla 10 donde se han ordenado de acuerdo al peor valor que puede arrojar el método. Este valor se calcula con  $\mu + \sigma = \mu_x + \sigma_x / 6$  y el mejor de ellos se calcula restando.

Tabla 10. Mejor y peor valor obtenido

Parámetros usados (Tabla 7)	Método	Peor Por $10^{281}$	Mejor Por $10^{281}$
G2	Kuri	1.51	1.43
G4	Kuri	1.65	1.41
G3	Kuri	1.70	1.46
G1	Kuri	1.70	1.43
G4	Powell	1.95	1.67
G4	Schoenauer	1.97	1.69
G3	Powell	2.11	1.84
G2	Powell	2.18	1.92
G1	Powell	2.18	1.88
G4	Joines	2.19	1.89
G4	Homaifar	2.33	1.92
G3	Joines	2.34	2.05
G2	Joines	2.36	2.09
G1	Joines	2.38	2.07
G1	Homaifar	2.40	2.08
G2	Homaifar	2.42	2.12
G3	Homaifar	2.45	2.19

Para facilitar la referencia de cada uno de los métodos analizados se usa una nomenclatura de 3 letras; las dos primeras indican los parámetros del experimento (G1 a G4) y la última es la inicial del autor del método. Así G4J refiere a los parámetros G4 con el método de Joines

Se puede apreciar que los métodos analizados (excepto Homaifar) aparecen juntos en la Tabla 10 y en la Figura 12 lo que nos da pie a suponer que éstos no dependen "demasiado" de los parámetros del algoritmo genético. Es decir, el mejor de ellos (Kuri) tiene este mismo comportamiento con cualesquiera parámetros del AG y el peor (Homaifar) mantiene su lugar frente a estos cambios.

Los datos mostrados nos sugieren que los parámetros seleccionados en G4 pueden hacer que el método sea más eficiente (caso de Homaifar, Joines y Powell) excepto en el método de Kuri. Sin embargo esta "superioridad" de G4 es más clara porque tanto G4 como G3 tienen poblaciones de 50 individuos. Lo que implica que hacen un menor número de evaluaciones de la función objetivo y obtienen información semejante o mejor que G1 y G2.

En la Figura 12 se han encerrado en elipses los resultados de cada uno de los métodos para enfatizar la formación de grupos y el mejor desempeño de los parámetros G4.

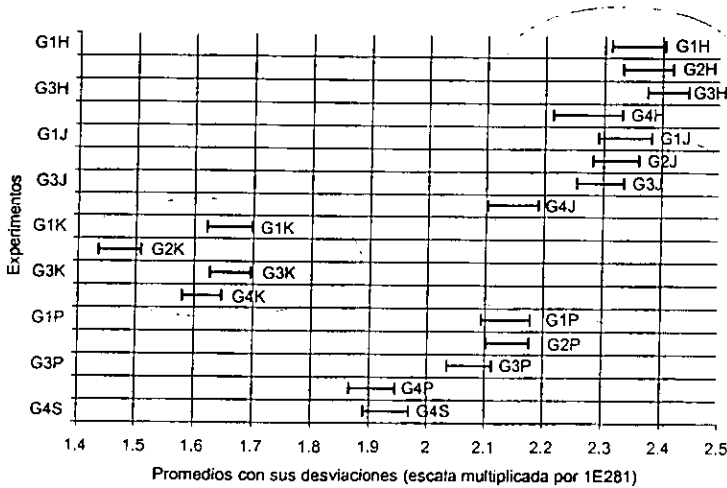


Figura 12. Promedios y desviaciones estándar de los experimentos realizados

Podría suponerse que los resultados presentados se han inducido al momento de realizar la normalización, porque el método utilizado para normalizar es muy similar al de Kuri. Pero para evitar esto se tomó una constante de normalización mucho mayor: en el método de Kuri la constante tiene un valor de  $10^9$  y el utilizado es  $5 \times 10^{281}$ .

Será necesario realizar un análisis posterior para cambiar este tipo de normalización a una similar a la de otros métodos (Joines por ejemplo) para tomar en cuenta no sólo el número de restricciones cumplidas sino también la magnitud de la restricción no cumplida.

## 5.4.1 Teorema de Chebyshev

La desviación estándar de una población es una medida de qué tan dispersos se encuentran los datos alrededor del promedio y es útil para comparar los resultados esperados al realizar una medición sobre los elementos de la muestra. El teorema de Chebyshev afirma lo siguiente:

“La proporción de cualquier distribución situada dentro de  $k$  desviaciones estándar de la media es por lo menos  $1 - \frac{1}{k^2}$ , donde  $k$  es cualquier número positivo mayor que 1”[16].

Utilizando este resultado y la Tabla 10 es posible expresar, por ejemplo, que al menos en un  $75\% = 0.75 = 1 - \frac{1}{2^2}$  de los experimentos realizados con el método de G2K será 1.59 veces mejor que G3H. El peor valor obtenido por G2K es menor que 1.56E251 y en caso de G3H menor que 2.49E281.

Este mismo cálculo se muestra en la Tabla 11. Si se desea conocer que tan bueno es un método respecto a otro sólo se debe buscar el punto de intersección.

Tabla 11. Comparativo de cada método

	G1H	G1J	G1K	G1P	G2H	G2J	G2K	G2P	G3H	G3J	G3K	G3P	G4H	G4J	G4K	G4P	G4S
G1H	1.0	1.0	1.4	1.1	1.0	1.0	1.6	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.5	1.2	1.2
G1J	1.0	1.0	1.4	1.1	1.0	1.0	1.6	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.4	1.2	1.2
G1K	0.7	0.7	1.0	0.8	0.7	0.7	1.1	0.8	0.7	0.7	1.0	0.8	0.7	0.8	1.0	0.9	0.9
G1P	0.9	0.9	1.3	1.0	0.9	0.9	1.4	1.0	0.9	0.9	1.3	1.0	0.9	1.0	1.3	1.1	1.1
G2H	1.0	1.0	1.4	1.1	1.0	1.0	1.6	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.5	1.2	1.2
G2J	1.0	1.0	1.4	1.1	1.0	1.0	1.6	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.4	1.2	1.2
G2K	0.6	0.6	0.9	0.7	0.6	0.6	1.0	0.7	0.6	0.6	0.9	0.7	0.6	0.7	0.9	0.8	0.8
G2P	0.9	0.9	1.3	1.0	0.9	0.9	1.4	1.0	0.9	0.9	1.3	1.0	0.9	1.0	1.3	1.1	1.1
G3H	1.0	1.0	1.4	1.1	1.0	1.0	1.6	1.1	1.0	1.0	1.4	1.2	1.0	1.1	1.5	1.3	1.2
G3J	1.0	1.0	1.4	1.1	1.0	1.0	1.5	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.4	1.2	1.2
G3K	0.7	0.7	1.0	0.8	0.7	0.7	1.1	0.8	0.7	0.7	1.0	0.8	0.7	0.8	1.0	0.9	0.9
G3P	0.9	0.9	1.2	1.0	0.9	0.9	1.4	1.0	0.9	0.9	1.2	1.0	0.9	1.0	1.3	1.1	1.1
G4H	1.0	1.0	1.4	1.1	1.0	1.0	1.5	1.1	1.0	1.0	1.4	1.1	1.0	1.1	1.4	1.2	1.2
G4J	0.9	0.9	1.3	1.0	0.9	0.9	1.5	1.0	0.9	0.9	1.3	1.0	0.9	1.0	1.3	1.1	1.1
G4K	0.7	0.7	1.0	0.8	0.7	0.7	1.1	0.8	0.7	0.7	1.0	0.8	0.7	0.8	1.0	0.8	0.8
G4P	0.8	0.8	1.1	0.9	0.8	0.8	1.3	0.9	0.8	0.8	1.1	0.9	0.8	0.9	1.2	1.0	1.0
G4S	0.8	0.8	1.2	0.9	0.8	0.8	1.3	0.9	0.8	0.8	1.2	0.9	0.8	0.9	1.2	1.0	1.0

Para construir la Tabla 11 se ha tomado el peor valor de cada método y se divide entre el peor correspondiente. Esto es G2P es 1.1 veces mejor que G3J en un 75% de los problemas que se resuelvan.

Un número mayor que uno (en el cruce) indica que el método en la columna es mejor que el que aparece en el renglón. En forma inversa si el número es menor que 1 y, finalmente si es 1, implica que tienen la misma eficiencia. Los datos se han redondeado a sólo un dígito decimal. Cada una de las entradas del triángulo superior de la matriz (Tabla 11) debe de ser el inverso de su correspondiente en el triángulo inferior (G4S es 1.2 veces mejor que G4H y G4H es 0.8 veces mejor que G4S porque  $1/1.2 = 0.83333334$ ).

Si se necesita una confiabilidad mayor sobre estos resultados, se deben calcular los peores valores (Tabla 10) sumando 4 desviaciones, lo que proporcionaría un valor de  $1 - \frac{1}{4^2} = .9375 = 93.75\%$ . En la Tabla 12 se ha calculado el cociente para distintos valores de  $k$ .

Tabla 12. Cocientes de desempeño para distintos intervalos

Valores de K	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6.0	7.0	8.0	9.0	10.0
G3H/G2K	1.6028	1.5919	1.5814	1.5712	1.5612	1.5516	1.5422	1.5331	1.5155	1.4989	1.4831	1.4680	1.4537
Porcentaje	55.56%	75%	84%	88.89%	91.84%	93.75%	95.06%	96.00%	97.22%	97.96%	98.44%	98.77%	99.00%

Es claro que de continuar incrementando el valor de  $k$  el cociente será cada vez menor. Pero el resultado importante es que en un 99% de los casos el mejor método supera al peor por 1.45. Esto significa que de 100 problemas resueltos con G2K y si se aplica G3H a estos mismos 45 de ellos no serán resueltos con una confianza del 99%.

Este resultado no supone que los datos se distribuyan en forma normal, pero sí se supone que la distribución de los datos (no de los promedios) es normal se sabe que en 3 desviaciones estándar está el 99% de los datos.

Una de las dudas importantes es: ¿Qué sucede con funciones que no se encuentran dentro de las 25 analizadas? Para responder se tomó la siguiente función

$$f(x) = \text{sen}(x_1) \cos(x_2) \quad (55)$$

$$-\pi < x_1, x_2 < \pi$$

tiene un máximo en  $f(x) = -1; x_1 = \pi/2, x_2 = \pi$ . Pero como  $x_2$  no puede tomar el valor de  $\pi$ , el resultado se debe aproximar a  $-1$  pero no llega a este valor. Se calculó su solución utilizando tanto el G2K, G2H y G3H. Obteniendo un los resultados de la Tabla 13.

Tabla 13. Resultados de tres métodos para una función distinta a las 25 analizadas

	Método (por 1E281)		
	G2K	G2H	G3H
Promedio	1.1015	1.683	2.116
Desviación	83.46	87.39	57.49
Peor valor	1.9361	2.5569	2.6909
Cociente	1.00	1.32	1.39

El problema se ejecutó 100 veces para cada método y de ahí se normalizó como se describió anteriormente. Aun cuando se mantiene el mejor desempeño del método G2K respecto a los otros el cociente es menor a todos los esperados (Tabla 12). Esto puede ser porque sólo se ha evaluado una función (en este caso un ejemplo sencillo) y el método de Homaifar no encuentra (en el caso de G3H) ningún punto factible. Pero, no es posible que los puntos que localice incumplan las cuatro restricciones y por ello no son castigados y la desviación es pequeña.

Para aclarar que el resultado anterior no está inducido por la posibilidad de que el G1K haya evaluado un número mayor de problemas sencillos (1-8) y un número menor de problemas complicados se realizó la gráfica de la Figura 13. En el eje horizontal se muestra el número de la función analizada y en el vertical la frecuencia con la que se evaluó. Esta frecuencia es respecto a la cantidad de experimentos que se realizaron (Tabla 8). Por ejemplo, si G1J tiene una frecuencia de 0.0312 para la función 8 significa que evaluó 101 veces la función ( $\sim 0.0312 \cdot 3240$ ).

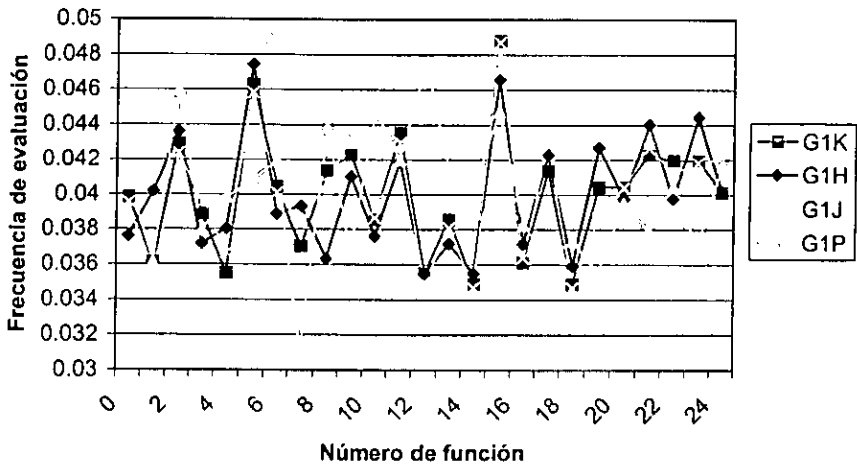


Figura 13. Frecuencia con la que se ejecuto un experimento para cada problema

En esta gráfica se muestra que ningún método fue favorecido para evaluar más funciones sencillas (las primeras) y menos funciones complicadas. Para los experimentos G2, G3 y G4 se tienen comportamientos similares

## **Capítulo 6. Conclusiones**

De los resultados presentados en la sección anterior la conclusión inmediata es que el método de Kuri tiene mejor desempeño que los restantes. Esto contradice la experiencia reportada respecto a estos métodos. En [28] se concluye lo siguiente:

"1) Penalties which are functions of the distance from feasibility are better performers than those which are merely functions of the number of violated constraints."

Esta contradicción puede estar generada por alguna de las siguientes causas(se describen sólo como hipótesis y sería necesario probar que son verdaderas).

- a) En este análisis comparativo se tomaron los mismos parámetros para todas las funciones y es necesario adecuar a cada problema el método que se utilice para seleccionar parámetros especiales.
- b) Es posible que el elitismo (que hace converger más rápido a una función) combinado con el GxK mantengan la población del AG muy variada y esto acelere el encontrar un mejor valor.
- c) La normalización que se realizó de cada resultado debe de contemplar la magnitud de la violación.
- d) También será necesario tomar en cuenta los "No Free Lunch Theorems" [32] para analizar si este resultado es una consecuencia de tales teoremas.

### **6.1 Trabajos Futuros**

Para mejorar el comparativo entre los distintos métodos propuestos será necesario tomar en cuenta qué tanto se han violado las restricciones. En este trabajo sólo se consideró si el método analizado había cumplido o no la restricción; pero el método de Joines puede, en algunos casos, proporcionar valores que violan "minimamente" las restricciones ( es más claro en las restricciones que tienen una igualdad a cero, donde el valor encontrado es pequeño pero no cero). Para realizar esto se debe cambiar la forma de normalizar las funciones y almacenar como parte de los resultados el vector solución encontrado por el método.

Hace falta también distinguir entre las restricciones con desigualdades y las que tienen igualdades para decir que método tiene mejor comportamiento en cada caso.

También será necesario comparar con otro tipo de métodos (no sólo función de penalización). En [2] se describen muchos otros que pueden ser implementados para realizar un análisis similar.



Por último, es deseable encontrar una forma de realizar la comparación sin la necesidad de conocer la solución del problema. Esto permitirá hacer el análisis no para un número finito de funciones (25 en este trabajo) sino utilizar una base de las funciones y las restricciones para poder general cualquier problema.

### Apéndice I. Descripción de las funciones de prueba.

En este apartado se describen las funciones de prueba que se han utilizado para hacer el análisis. Estas funciones han sido tomadas de los artículos que describen los métodos analizados y de varios libros más. Cabe mencionar que muchas de las funciones que se usan en un artículo aparecen en otros o en los apéndices de libros que las proporcionan como funciones complicadas.

Para seleccionar la muestra de funciones a utilizar se eligieron 5 que fueran muy sencillas y que aparecen en los libros de programación lineal básica; las siguientes 5 son funciones que expresan una situación del mundo real (ver la referencia para detalles) las siguientes 10 son funciones citadas en distintos artículos que tratan optimización restringida con AG y las 5 últimas son combinaciones de funciones trigonométricas.

Como los dominios de cada función son distintos (por ejemplo  $\mathfrak{R}, \mathfrak{R}^2, \dots, \mathfrak{R}^{10}$ ) fue necesario trazar un hipercubo en el cual estuvieran contenidos todos ellos. Este está dado por los parámetros del AG (en todos se utilizaron 20 dígitos binarios para la parte decimal y 14 para la parte entera). Con esta región se calculó la proporción entre ésta y la región factible para cada función:

$$\rho = \frac{|M|}{|S|} \quad (56)$$

donde  $M$  es el conjunto de los puntos que satisfacen las restricciones y  $S$  el rectángulo en  $\mathfrak{R}^n$  que contiene a todas las regiones factibles de las funciones analizadas en este trabajo. Para calcular  $\rho$  se generaron 10'000,000 de puntos en  $S \subset \mathfrak{R}^n$  y se revisó si cumplían con las restricciones o no.

Los valores calculados para cada una de las funciones aparecen en la Tabla 14.

Tabla 14. Valores de  $\rho$  para cada función

Función	Proporción $\rho$	Función	Proporción $\rho$
1	50.0116	14	0.0
2	49.9729	15	0.0
3	0.0	16	99.9515
4	66.9322	17	0.0
5	0.0	18	0.0
6	0.0	19	0.0
7	0.0	20	0.0
8	0.0	21	0.0
9	0.0	22	0.0
10	0.0	23	19.9414
11	0.0	24	0.9307
12	0.0	25	0.0
13	0.0		

## ***1.1 Funciones***

En la tabla de funciones que se muestra a continuación aparece lo siguiente: la primera columna muestra la referencia donde se obtuvo la función y el número con que asignado para este trabajo; en la segunda, la expresión de la función objetivo; la tercera tiene las restricciones y la cuarta la solución con el mejor valor conocido, (el cual es el óptimo global en algunos casos).

Finalmente se graficaron algunas de estas funciones (sólo las que están en dos y tres dimensiones) para poder apreciar su complejidad y su región factible.

Número	Función	Restricciones	Solución
1	$0.5x_1$	$x_1 - 1 \geq 0$	$f(x) = 0.5$ $x_1 = 1$
2	$x_1^{0.5}$	$\frac{1}{4}x_1 \geq 1$	$f(x) = 2$ $x_1 = 4$
3	$x_1^3 - 6x_1^2 + 11x_1 + x_3$	$x_1^2 + x_2^2 - x_3^2 \leq 0$ $x_1^2 + x_2^2 + x_3^2 \geq 4$ $x_3 \leq 5; x_1 \geq 0; x_2 \geq 0; x_3 \geq 0$	$f(x) = \sqrt{2}$ $x_1 = 0; x_2 = \sqrt{2};$ $x_3 = \sqrt{2}$
4	$(x_1 - x_2)^2 + (x_2 - x_3)^4 + 1$	$x_1 + x_1x_2^2 + x_3^4 = 3$	$f(x) = 1$ $x_1 = x_2 = x_3 = 1$
5	$x_1^2 + 4x_2^2$	$\frac{3}{5}x_1 + \frac{4}{5}x_2 \geq \frac{13}{5}$	$f(x) = 13$ $x_1 = 3; x_2 = 1$
6	$9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 + x_3^2$ $+ 2x_1x_2 + 2x_1x_3$	$x_1 \geq 0; x_2 \geq 0; x_3 \geq 0$ $-x_1 - x_2 - x_3 \geq -3$	$f(x) = 3/27$ $x_1 = 4/3; x_2 = 7/9;$ $x_3 = 4/9$
7	$(x_1 - 2)^2 + (x_2 - 1)^2$	$x_1 - 2x_2 + 1 = 0$ $-\frac{x_1^2}{4} - x_2^2 + 1 \geq 0$	$f(x) = 1.393$ $x_1 = 0.823$ $x_2 = 0.911$
8	$1000 - x_1^2 - 2x_2^2 - x_3^2 - x_1x_2 - x_1x_3$	$x_1^2 + x_2^2 + x_3^2 - 25 = 0$ $8x_1 + 14x_2 + 7x_3 - 56 = 0$ $x_i \geq 0; i = 1, 2, 3$	$f(x) = 961.715$ $x_1 = 3.512; x_2 = 0.217;$ $x_3 = 3.552$ $h_1(x) = h_2(x) = 0$
9	$100(x_2 - x_1^2) + (1 - x_1)^2 + 90(x_4 - x_3^2)^2$ $+ (1 - x_3)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2]$ $+ 19.8(x_2 - 1)(x_4 - 1) + 1$	$-10 \leq x_i \leq 10$ $i = 1, 2, 3, 4$	$f(x) = 1$ $x_1 = 1; x_2 = 1;$ $x_3 = 1; x_4 = 1$

Número	Función	Restricciones	Solución
10 [11]	$5.3578547x_3^2 + 0.8356891x_1x_5$ $+ 37.293239x_1 - 40792.141$	$0 \leq 85.334407 + 0.0056858x_2x_3$ $+ 0.0006262x_1x_4 - 0.0022053x_3x_5 \leq 92$ $90 \leq 80.51249 + 0.0071317x_2x_3$ $+ 0.0029955x_1x_2 + 0.0021813x_3^2 \leq 110$ $20 \leq 9.300961 + 0.0047026x_3x_5$ $+ 0.0012547x_1x_3 + 0.0019085x_3x_4 \leq 25$ $78 \leq x_1 \leq 102; 33 \leq x_2 \leq 45; 27 \leq x_3 \leq 45$ $27 \leq x_4 \leq 45; 27 \leq x_5 \leq 45$	$f(x) = -30367$ $x_1 = 78.62$ $x_2 = 33.44$ $x_3 = 31.07$ $x_4 = 44.18$ $x_5 = 35.22$
11 [25]	$5x_1 + 5x_2 + 5x_3 + 5x_4 - 5 \sum_{i=1}^4 x_i^2 - \sum_{i=2}^{13} x_i$	$2x_1 + 2x_2 + x_{10} + x_{11} \leq 10$ $2x_1 + 2x_3 + x_{10} + x_{12} \leq 10$ $2x_2 + 2x_3 + x_{11} + x_{12} \leq 10$ $-8x_1 + x_{12} \leq 0$ $-2x_4 - x_3 + x_{10} \leq 0$ $-2x_6 - x_7 + x_{11} \leq 0$ $-2x_8 + x_9 + x_{12} \leq 0$ $0 \leq x_i \leq 1; i = 1, \dots, 9$ $0 \leq x_i \leq 100; i = 10, 11, 12$ $0 \leq x_{13} \leq 1$	$f(x) = -15$ $x_i = 1; i = 1, \dots, 9$ $x_{10} = x_{11} = x_{12} = 3$ $x_{13} = 1$

Número	Función	Restricciones	Solución
12 [25]	$x_1 + x_2 + x_3$	$1 - 0.0025(x_4 + x_6) \geq 0$ $1 - 0.0025(x_5 + x_7 - x_4) \geq 0$ $1 - 0.01(x_8 - x_5) \geq 0$ $x_1 x_6 - 833.33252x_4 - 100x_1 + 833333.333 \geq 0$ $x_2 x_7 - 1250x_5 - x_2 x_4 + 1250x_4 \geq 0$ $x_3 x_8 - 1250000 - x_3 x_5 + 2500x_5 \geq 0$ $100 \leq x_1 \leq 10,000; 1,000 \leq x_i \leq 10,000; i = 2,3$ $10 \leq x_j \leq 1,000; j = 4, \dots, 8$	$f(x) = 7049.330923$ $x_1 = 579.3167;$ $x_2 = 1359.943$ $x_3 = 5110.071;$ $x_4 = 182.0174$ $x_5 = 295.5985;$ $x_6 = 217.9799$ $x_7 = 286.4162;$ $x_8 = 395.5979$
13 [25]	$(x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2$ $+ 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6 x_7 - 10x_6 - 8x_7$	$127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0$ $282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0$ $196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0$ $-4x_1^2 - x_2^2 + 3x_1 x_2 - 2x_3^3 - 5x_6 + 11x_7 \geq 0$ $-10.0 \leq x_i \leq 10.0; i = 1, \dots, 7$	$f(x) = 680.6300573$ $x_1 = 2.330499;$ $x_2 = 1.951372;$ $x_3 = -0.4775414$ $x_4 = 4.365726;$ $x_5 = -0.6244870;$ $x_6 = 1.038131$ $x_7 = 1.594227$
14 [25]	$e^{x_1 x_2 x_3 x_4 x_5}$	$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 10$ $x_1 x_3 - 5x_4 x_5 = 0$ $x_1^3 + x_2^3 = -1$ $-2.3 \leq x_i \leq 2.3; i = 1, 2$ $-3.2 \leq x_j \leq 3.2; j = 3, 4, 5$	$f(x) = 0.0539498478$ $x_1 = -1.717143;$ $x_2 = 1.595709$ $x_3 = 1.827247;$ $x_4 = -0.7636413$ $x_5 = -0.7636450$

Número	Función	Restricciones	Solución
15 [25]	$x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2$ $+ 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2$ $+ 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2$ $+ (x_{10} - 7)^2 + 45$	$105 - 4x_1 - 5x_2 + 3x_7 - 9x_8 \geq 0$ $-10x_1 + 8x_2 + 17x_3 - 2x_8 \geq 0$ $8x_1 - 2x_2 - 5x_9 + 2x_{10} + 12 \geq 0$ $-3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + 7x_4 + 120 \geq 0$ $-5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0$ $-x_5^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_3 + 6x_6 \geq 0$ $-0.5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_3^2 + x_6 + 30 \geq 0$ $3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0$ $-10.0 \leq x_i \leq 10.0; i = 1, \dots, 10$	$f(x) = 24.3062091$ $x_1 = 2.171996;$ $x_2 = 2.363683;$ $x_3 = 8.773926$ $x_4 = 5.095984;$ $x_5 = 0.9906548;$ $x_6 = 1.430574$ $x_7 = 1.321644;$ $x_8 = 9.828726;$ $x_9 = 8.280092$ $x_{10} = 8.375927$
16 [20]	$3x_1^2 + 2 \operatorname{sen}(x_1) \cos(x_2) + 2x_2^2$ $+ \cos(x_1) \operatorname{sen}(2x_2) - 27.5$	$3x_1^2 + 2 \operatorname{sen}(x_1) \cos(x_2) - 10.2 \geq 0$ $2x_2^2 + \cos(x_1) \operatorname{sen}(2x_2) - 17.3 \geq 0$	$f(x) \approx -1.0037418244$ $x_1 \approx 1.9981788993$ $x_2 \approx 2.9264232963$
17 [20]	$-\{x_1 +  \operatorname{sen}(32x_1)  -  \cos(11x_2) \}$	$0 \leq x_1 \leq \pi \quad i = 1, 2$	$f(x) \approx 4.0930$ $x_1 \approx 3.0935; x_2 \approx 2.1420$
18 [18]	$3x_1 + 0.000001x_1^3 + 2x_2 + \frac{0.000002}{3}x_2^3$	$x_4 - x_3 + 0.55 \geq 0; x_3 - x_4 + 0.55 \geq 0$ $1000 \operatorname{sen}(-x_3 - 0.25) + 1000 \operatorname{sen}(-x_4 - 0.25)$ $+ 894.8 - x_1 = 0$ $1000 \operatorname{sen}(x_3 - 0.25) + 1000 \operatorname{sen}(x_3 - x_4 - 0.25)$ $+ 894.8 - x_2 = 0$ $1000 \operatorname{sen}(x_4 - 0.25) + 1000 \operatorname{sen}(x_4 - x_3 - 0.25)$ $+ 1294.8 = 0$ $0 \leq x_i \leq 1200; i = 1, 2; -0.55 \leq x_i \leq 0.55; i = 3, 4$	$f(x) = 5126.4981$ $x_1 = 679.9453$ $x_2 = 1026.067$ $x_3 = 0.1188764$ $x_4 = -0.3962336$

Número	Función	Restricciones	Solución
19 [18]	$-\frac{\text{sen}^3(2\pi x_1) \text{sen}(2\pi x_2)}{x_1^3(x_1 + x_2)}$	$0 \leq x_i \leq 10; i = 1, 2$ $x_1^2 - x_2 + 1 \leq 0$ $1 - x_1 + (x_2 - 4)^2 \leq 0$	$f(x) = -0.09582503$ $x_1 = 1.2279348$ $x_2 = 4.2453756$
20 [11]	$\sum_{i=1}^{10} \{ \ln(x_i - 2) \}^2 + \left[ \ln(10 - x_i) \right] - \left( \prod_{i=1}^{10} x_i \right)^{0.2}$	$2.001 < x_i < 9.999$ $i = 1, \dots, 10$	$f(x) = -45.778$ $x_i = 9.351; i = 1, \dots, 10$
21 [13]	$f_1(x_1) + f_2(x_2)$ $f_1(x) = \begin{cases} 30x & 0 \leq x < 300 \\ 31x & 300 \leq x < 400 \end{cases}$ $f_2(x) = \begin{cases} 28x & 0 \leq x < 100 \\ 29x & 100 \leq x < 200 \\ 30x & 200 \leq x < 1000 \end{cases}$	$x_1 = 300 - \frac{x_3 x_4}{131.078} \cos(1.48477 - x_6)$ $+ \frac{0.90798 x_3^2}{131.078} \cos(1.47588)$ $x_2 = -\frac{x_3 x_4}{131.078} \cos(1.48477 + x_6)$ $+ \frac{0.90798 x_4^2}{131.078} \cos(1.47588)$ $x_3 = -\frac{x_3 x_4}{131.078} \text{sen}(1.48477 + x_6)$ $+ \frac{0.90798 x_4^2}{131.078} \text{sen}(1.47588)$ $200 - \frac{x_3 x_4}{131.078} \text{sen}(1.48477 - x_6)$ $+ \frac{0.90798 x_3^2}{131.078} \text{sen}(1.47588) = 0$ $0 \leq x_1 \leq 400; 0 \leq x_2 \leq 1000; 340 \leq x_3 \leq 420$ $340 \leq x_4 \leq 420; -1000 \leq x_5 \leq 1000$ $0 \leq x_6 \leq 0.5236$	$f(x) = 8927.5888$ $x_1 = 107.81$ $x_2 = 196.32$ $x_3 = 373.83$ $x_4 = 420.00$ $x_5 = 21.31$ $x_6 = 0.153$
22 [81]	$-(e^{-x_1} + x_1 + (x_2 + 1)^2)$	$x_1^2 + x_2 \leq 3$ $x_1 \geq 0; x_2 \geq 0$	$f(x) = -17$ $x_1 = 0; x_2 = 3$



Número	Función	Restricciones	Solución
23	$\begin{aligned} & \text{sen}(x_1, x_2) \cos(x_3, x_4) + \text{sen}(x_3, x_4) \cos(x_1, x_2) \\ & + 2x_1 \text{sen}(\pi x_3) + x_3 \cos(\pi x_1) \\ & + \sum_{i=1}^4 \text{sen}(\pi x_i) + 4x_1^2 - x_2 + x_4 \\ & - 6.807418876198037 \end{aligned}$	$\begin{aligned} & \text{sen}(x_1, x_2) \cos(x_3, x_4) + \text{sen}(x_3, x_4) \cos(x_1, x_2) \\ & + 0.807418876 \geq 0 \\ & \sum_{i=1}^4 \text{sen}(\pi x_i) \geq 0 \\ & 4x_1^2 - x_2 + x_4 - 5 \geq 0 \\ & 2x_1 \text{sen}(\pi x_3) + x_3 \cos(\pi x_1) - 1 \geq 0 \end{aligned}$	$\begin{aligned} f(\vec{x}) &= 1 \\ x_1 &= 0.5 \\ x_2 &= 3.5 \\ x_3 &= 0.5 \\ x_4 &= 7.5 \end{aligned}$
24	$\begin{aligned} & e^{x_1} \cos(x_2) + e^{x_3} \cos(x_1) + \ln(x_1) + x_2^2 \\ & + 1.12330837 \end{aligned}$	$\begin{aligned} & e^{x_1} \cos(x_2) + e^{x_3} \cos(x_1) - 2.18712491 \geq 0 \\ & \ln(x_1) + x_2^2 + 2.063816542 \geq 0 \end{aligned}$	$\begin{aligned} f(\vec{x}) &= 1 \\ x_1 &= 1/16 \\ x_2 &= 1/8 \end{aligned}$
25	$\begin{aligned} & (x_1^2 + x_2^2) \times  \text{sen}(10(x_1^2 + x_2^2))  \\ & \times  \cos(10(x_1^2 + x_2^2))  + (x_1^2 + x_2^2 + 3)/2 \end{aligned}$	$\begin{aligned} & x_1 + x_2^2 + 1 \geq 0 \\ & x_1^2 + x_2 + 1 \geq 0 \\ & -2 \leq x_1 \leq 2, -2 \leq x_2 \leq 2 \end{aligned}$	$\begin{aligned} f(\vec{x}) &= 1.5 \\ x_1 &= 0 \\ x_2 &= 0 \end{aligned}$

## 1.2 Gráficas

La gráfica de cada una de las funciones descritas en la tabla anterior se muestra en las siguientes figuras. No aparecen las 25 puesto que algunas de ellas no se pueden visualizar en 3 dimensiones. Los números con que se hace referencia en este caso corresponden a los de la tabla del Apéndice I. En los casos en que se considera interesante también se ha graficado el dominio de la función.

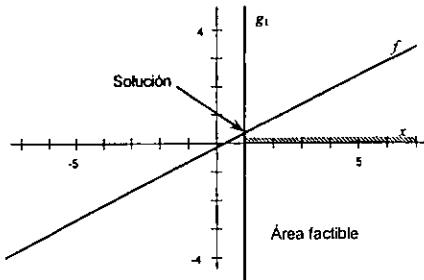


Figura 14. Función 1 con área factible

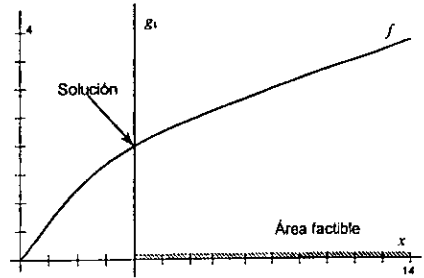


Figura 15. Función 2 con área factible

La función 1 y 2 son de  $\mathbb{R} \rightarrow \mathbb{R}$ , su solución es muy simple y se ha mostrado el dominio de la función junto con su gráfica.

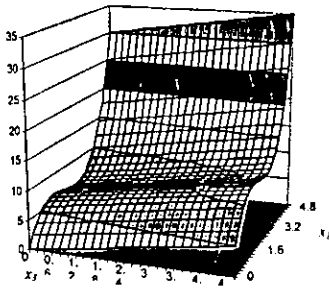


Figura 16. Función 3

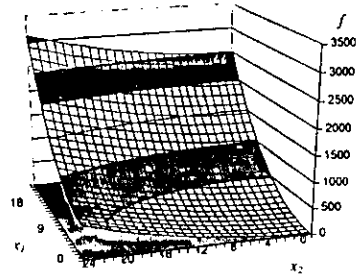


Figura 17. Función 5

El dominio de la función 3 no se puede graficar porque tiene tres variables independientes. La función 5 es suave y continua el punto mínimo se muestra en Figura 18 con la gráfica de la restricción.

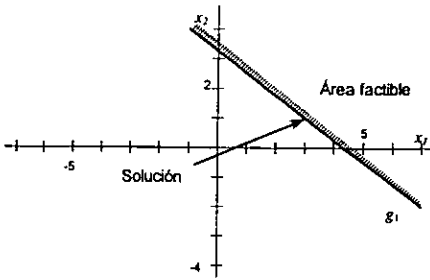


Figura 18. Región factible de la función 5

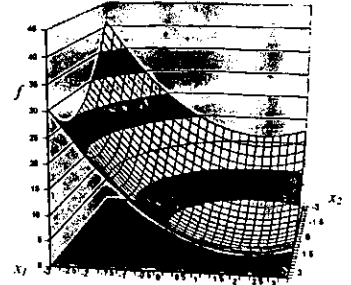


Figura 19. Función 7

La función 7 es un paraboloides y sólo se muestra el cuadrante que contiene a la región factible del problema.

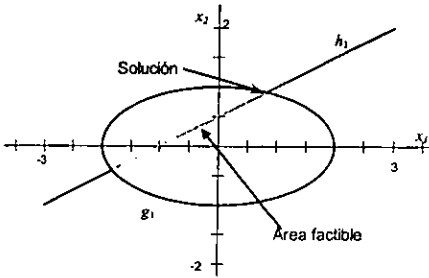


Figura 20. Región factible función 7

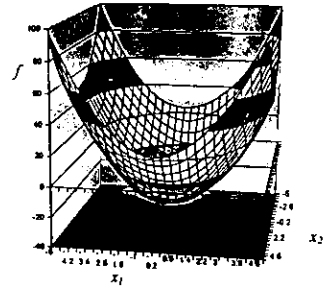


Figura 21. Gráfica de la función 16

La función 16 es simple y sus restricciones son funciones casi rectas pero que tienen un comportamiento asintótico. La zona no factible forma una cruz.

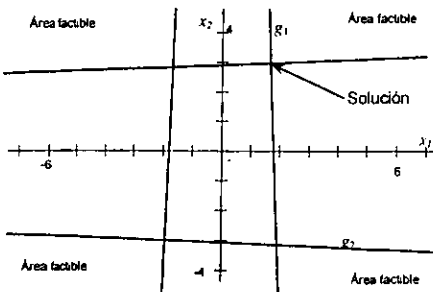


Figura 22. Restricciones función 16

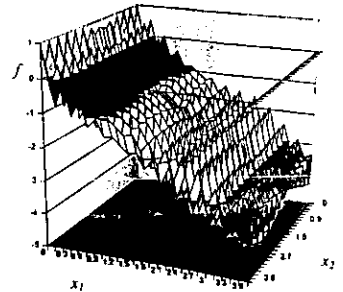


Figura 23. Gráfica de la función 17

La función 17 (Figura 23) tiene muchos picos y su valor decrece al acercarse a la frontera de la región factible (no se graficó porque sólo es un cuadro de lado  $\pi$  y una de sus esquinas está en el origen).

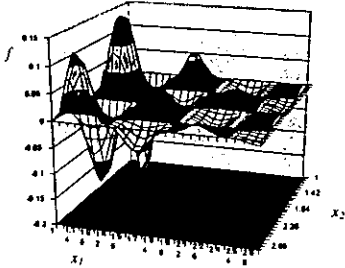


Figura 24. Gráfica de la función 19

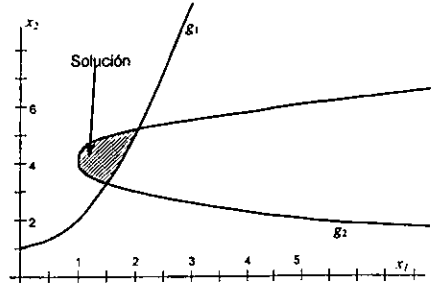


Figura 25. Restricciones de la función 19

La función 19 tiene forma de ondas y la región factible es simple pero relativamente pequeña para todo el dominio que se está considerando.

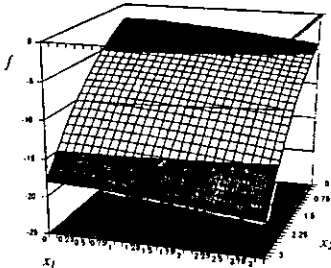


Figura 26. Gráfica función 22

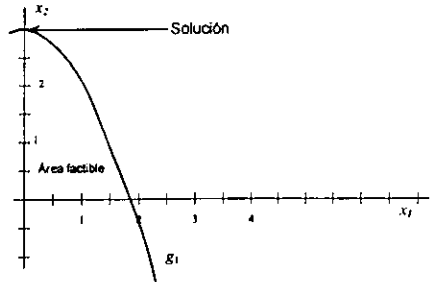


Figura 27. Restricciones de la función 22

En la función 22 (Figura 27) la solución se encuentra en la orilla de la región factible. En la Figura 26 es posible ver que hay valores menores pero estos ya no satisfacen la restricción.

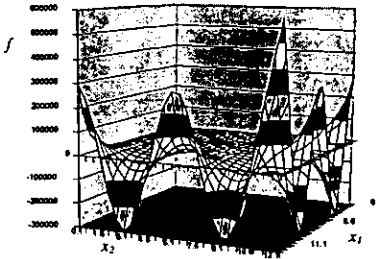


Figura 28. Gráfica función 24

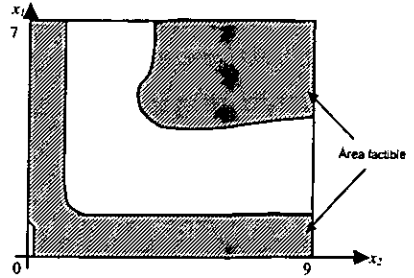


Figura 29. Restricciones de la función 24

Esta función 24 (Figura 28) no tiene una región factible, sino "varias zonas". Si se gráfica el dominio para una escala mayor se puede ver que se alternan áreas factibles y no factibles (Figura 29).

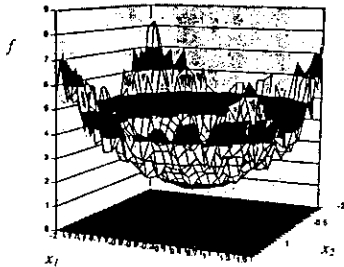


Figura 30. Gráfica de la función 25

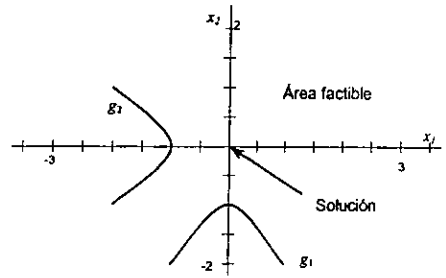


Figura 31. Restricciones de la función 25

El número de picos en la función 25 (Figura 30) es causado por el valor absoluto de la función. El mínimo se encuentra en el centro de la figura y coincide con el origen.

## Apéndice II. Descripción del software utilizado.

El software desarrollado para realizar los experimentos descritos en esta tesis se construyó utilizando *Delphi 4.0* [23]. Se implementó el AG utilizando dos clases: genoma y población. La primera de ellas representa a cada genoma como un arreglo de palabras (tipo *LongWord*) y se utilizan *bits* para representar a cada uno de los elementos de la cadena del genoma. En la clase población se usa una lista dinámica (*Tlist*) para almacenar a los individuos. El comportamiento que se definió para cada clase se muestra en la Tabla 15.

Tabla 15. Definición de las clases utilizadas en la implementación

Genoma	Comentario
GenToStr	Convierte un genoma a cadenas de ceros y unos
ObtenerEntero	Obtiene los bits que forman a un entero dentro de todo el genoma
ObtenerExtended	Obtiene los bits que forman a un número real dentro de todo el genoma
GrayToBin	Convierte un genoma en binario pesado a Gray
Variable	Obtiene el valor de una variable del genoma
Mutar	Implementa la mutación en un genoma
CruzaPf	Implementa la cruza en punto fijo
Cruza2p	Implementa la cruza en dos puntos fijos
CruzaUn	Implementa la cruza uniforme
Copiar	Copia un genoma en otro
Población	
Suma	Regresa la suma de las aptitudes de la población
EligeIndRul	Implementa la selección de ruleta
EligeIndVas	Implementa la selección de Vasconcelos
EligeIndNie	Implementa la selección de Nietzsche
DeathPenalty	Implementa método de Kuri
StaticPenalty	Implementa método de Homaifar
DynamicPenalty	Implementa método de Joines
ShoenauerPenalty	Implementa método de Shoenauer
PowellPenalty	Implementa método de Powell
Despliega	Muestra la población en un <i>Tgrid</i>
Promedio	Calcula el promedio de las aptitudes de población
Varianza	Calcula la varianza de las aptitudes de población
Ordena	Ordena de menor a mayor la población de acuerdo a la aptitud
Calificar	Asigna a cada individuo de la población el valor de la aptitud correspondiente
SigGenera	Calcula la generación siguiente a partir de la actual
Mejores	Calcula los datos del mejor individuo de la población actual

Después de esto, se construyó una interfaz MDI (Interfaz de documento múltiple) en donde cada ventana hija corresponde a un experimento determinado. Se

agregó a cada ventana hija un menú de árbol. El procedimiento más sencillo para utilizar el programa es el siguiente:

- Ejecutar el programa.
- Seleccionar **A**rchivo | **N**uevo.
- Seleccionar Problema en el menú de árbol de la ventana documento.
- Presionar el botón **F**unción.
- Aparecerá un cuadro de diálogo donde podrá editarse la función a optimizar. Teclee la función en el recuadro superior y las restricciones en la lista izquierda. Es necesario especificar primero el nombre de las variables que se utilizarán en la lista de la derecha. En el mismo diálogo aparece un texto de ayuda. Puede guardar su función en un archivo o cargar una anteriormente utilizada con los botones de **G**uardar y **T**raer de Archivo respectivamente. Al terminar presione el botón aceptar. Si existe algún error aparecerá un aviso; tendrá que presionar aceptar y corregir el error.
- Después podrá especificar los parámetros del AG. Sólo debe presionar en el árbol la opción que desea modificar y en la parte derecha aparecerán cajas de edición con los valores correspondientes.

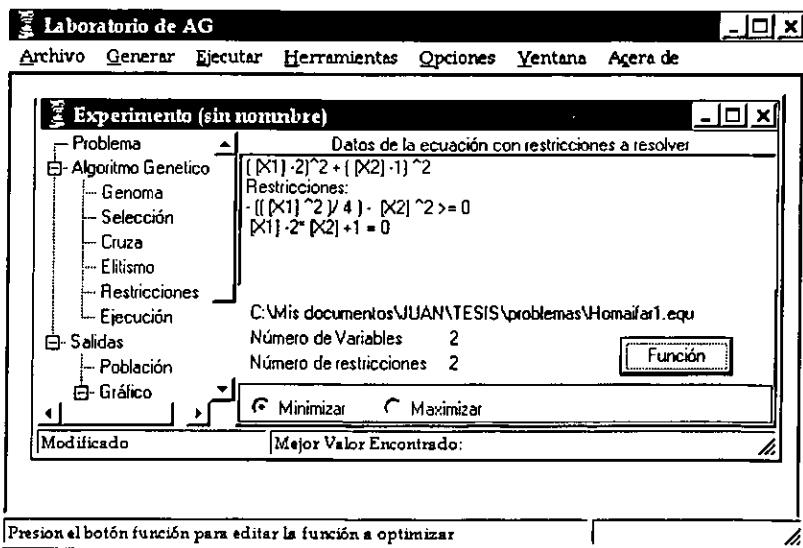


Figura 32. Ventanas del programa implementado para realizar los experimentos descritos. Se muestra una ventana documento.

- Para ejecutar el AG se selecciona la pestaña de "Ejecución", o el mismo en el menú de árbol, el botón "Crear" y después especificar el número de generaciones (100 por omisión) y presionar el botón "Ejecutar 100 generaciones".
- En la barra de estado de la ventana documento aparecerá el número de generaciones ejecutadas hasta el momento y el mejor valor encontrado.

- Si desea, puede presionar en el menú de árbol la opción de “gráfica” para ver el promedio de toda la población, la varianza, el mejor valor encontrado por generación o el número de restricciones cumplidas por generación.

Al terminar de ejecutar todas las generaciones que se le han especificado (o antes si presiona el botón de Detener que se encuentra en la opción de ejecución del menú de árbol) se puede ver la población completa en la opción Salidas población. Si desea guardar estos datos para análisis posteriores puede presionar el botón derecho del ratón sobre la población y escoger “guardar” o “copiar”.

También se implementó un modo de ejecución por lotes (*batch*). Ya sea en una ventana o en varias (la opción generar del menú de la ventana de aplicación). Se pueden ejecutar varios experimentos en serie. Esta fue la opción utilizada para obtener los resultados descritos.

La ejecución en varias ventanas sólo permite 20 experimentos distintos (el número depende de la configuración de *Windows* y cuántas ventanas hijas permita a una aplicación) pero se pueden almacenar todos los datos generados (las gráficas y la población completa). En cambio, en la ejecución en la misma ventana (no hay límite para el número de experimentos a ejecutar) sólo se almacenan los mejores valores obtenidos para cada experimento, el vector solución y el número de restricciones que cumple este vector solución.

La ejecución en la misma ventana permite cambiar de problema (función objetivo); selecciona un problema al azar de un conjunto previamente descrito (las 25 funciones descritas en el anexo 1 en este caso).

Para probar que el generador de números aleatorios utilizado (que está integrado con el lenguaje) tiene una distribución uniforme se realizó la prueba de frecuencias (*frequency test*) que se describe en [19].



### ***Apéndice III. Aptitud vs. Generación***

Con el software desarrollado, como parte de esta tesis, se puede apreciar en una gráfica el valor de la aptitud del mejor individuo en cada generación. Se esperaría que estas gráficas sean siempre decrecientes en caso de utilizar elitismo, pero esto no es así porque en algunos métodos el mejor individuo de la generación  $i$  no se le asigna el mismo valor de la aptitud en la generación  $i+1$ . De los métodos comparados en el capítulo 4 se describen a continuación cuales de ellos presentan este comportamiento y sus causas.

En el caso del método de Joines, la dependencia de la función de penalización sobre el número de generación está explícita en su definición (41). Esta dependencia hará que la aptitud crezca como  $Ck^a$  mientras no se encuentre un mejor individuo de una generación a otra. En la gráfica de aptitud vs. generación no es posible apreciar cuando se satisface una restricción más. Porque aún cuando al encontrar un punto que satisfacen más condiciones se provoca un brinco (cambia de creciente a decreciente) no todos estos cambios están generados por satisfacer una restricción más. Porque el AG está buscando minimizar también las restricciones y es posible encontrar un valor menor para las restricciones y que aún no sean cumplidas.

En el método de Schoenauer se cambia la función de aptitud. Los individuos de las primeras generaciones se califican de acuerdo al número de restricciones que les falta por cumplir hasta que un cierto porcentaje de la población cumpla todas las restricciones y en la generación donde esto se cumpla se comienza a calificar a los individuos con la función a minimizar. Este cambio puede provocar el mismo efecto de brinco que se describió anteriormente. Pero para este método solamente aparece una vez.

Este comportamiento en los AG elitistas también se presenta en la gráfica del número de restricciones cumplidas vs. generación. En una generación  $i$  pueden el mejor individuo satisfacer  $j$  restricciones y en la generación  $i+1$  el mejor individuo satisface menos restricciones. Esto estará ocasionado porque la función de penalización cae en el problema que se describió para el algoritmo de Homaifar, es decir, individuos que cumplen menos restricciones tienen una calificación menor a la asignada a individuos que satisfacen más restricciones. Sin embargo los algoritmos de penalización dinámica pueden recuperar el número de restricciones cumplidas.

En el caso de la función de penalización de Kuri no presenta ninguno de los comportamientos anteriores; esto está implícito en que no tiene una dependencia explícita de la magnitud de la violación, sino sólo del número de restricciones no cumplidas.

**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**

**Bibliografía y Referencias.**

- [1] Avriel, M. "Nonlinear programming: Analysis and Methods" Prentice may. 1976. pp 512.
- [2] Coello C. "Constraint-handling techniques used with evolutionary algorithms: A survey of the state of the art" Computer Methods in Applied Mechanics and Engineering. 2001 (aceptado para publicación).
- [3] Dennis, J. "Mathematical Programs and Electrical Networks" John Wiley & Sons, Inc. New York (1959).
- [4] Galaviz, J. And Kuri, A "A self-Adaptative Genetic Algorithm for Funtion Optimization", Proccedings ISAI/IFIPS. pp 156. (1996)
- [5] Gass S. "Linear Programming Methods and Applications" McGraw Hill. 1975.
- [6] Gen M. and Cheng R. "A Survey of Penalty Techniques in Genetic Algorithms". 1996.
- [7] Goldberg, D. "Genetic Algorithms in Search, Optimization, and Machine Learning". Addison Wesley Publishing Company, Inc. 1989
- [8] Guenther W. "Introducción a la inferencia estadística". 1968. 357 págs.
- [9] Hajela, P. "Genetic Search – An Approach to the Nonconvex Optimization Problem" AIAA Journal. Vol. 26, No 7, 1990, pp 30-39.
- [10] Hinterding R, Michalewicz Z & Eiben A.. "Adaptation in Evolutionary Computation: A Survey" Proceedings of the 4th IEEE International Conference on Evolutionary Computation, Indianapolis, Abril 1997, pp.65-69.
- [11] Himmelblau, D. "Applied Nonlinear Programming". Mc Graw Hill, (1972).
- [12] Holland, J. "Adaptation in Natural and Artificial System". Ann Arbor, The University of Michigan Press, (1975).
- [13] Homaifar A., Qi C. & Lai S. "Constrained Optimization Via Genetic Algorithms". Simulation 62:4, 242-254. 1994.
- [14] Hooke, R y Jeeves T "Assoc. Computer Mach" 8:121 (1962).
- [15] Jobson, J. "Applied Multivariate Data Analysis" Springer – Verlag New York. 1991. volumen 1. 621 págs.
- [16] Johnson, R. "Estadística elemental". Trillas. 1976. pp 515
- [17] Joines, J and Houck, C. "On the Use of Non-Stationary Penalty Functions to Solve Nonlinear Constrained Optimization Problems with GA's". Proceedings of first IEEE Conference on Evolutionary Computation. pp 579-584. 1994.
- [18] Koziel, S. and Michalewicz, Z. "Evolutionary Algorithms, Homomorphous Mappings, and Constrained Parameter Optimization". 1999.
- [19] Kuri, A. "A Comprehensive Approach to Genetic Algorithms in Optimization and Learning. Theory and Applications". Volumen 1. Colección de Ciencias de la Computación. Instituto Politécnico Nacional. (1999)
- [20] Kuri, A. "A Statical Genetic Algorithm". Memorias del ENC99, Pachuca México 1999. 102 págs.
- [21] Kuri, A. "A universal Eclectic Genetic Algorithm for Constrained Optimization". Proccedings 6<sup>th</sup> European Congress on Intelligent Techniques & Soft Computing, EUFIT'98. pp 518-522. 1998.
- [22] Marsden, J. "Cálculo vectorial" Fondo Educativo Interamericano. (1982).

- [23] Marteens, I. "La cara oculta de Delphi 4 programación avanzada para bases de Datos". Daysoft Internacional. 1998. 912 págs.
- [24] Michalewicz Z. "A Survey of Constraint Handling Techniques in Evolutionary Computation Methods". Proceedings of the 4<sup>th</sup> Annual Conference on Evolutionary Programming. MIT Press. Pág. 135 – 155.1995.
- [25] Michalewicz, Z. "Genetic Algorithms + Data Structures = Evolution Programs". 3ª edición. Springer (1996). pp 215
- [26] NIST Statistical Engineering Division WWW Home Page. <http://www.nist.gov/itl/div898/>
- [27] Powell, D. and Skolnick, M. "Using Genetic Algorithms in Engineering Design Optimization with Non-linear Constraints". Proceedings of the Fifth International Conference on Genetic Algorithms. 1993. pp 424 430.
- [28] Richardson J., Palmer M., Liepins G. & Hilliard M. "Some Guidelines for Genetic Algorithms with Penalty Functions" Proceedings of the IEEE International Conference on Evolutionary Computation, 1989, pp.191-197.
- [29] Shoenauer, M. and Xanthakis, S. "Constrained GA Optimization". Proceedings of the Fifth International Conference on Genetic Algorithms. 1993 pp 573-580
- [30] Snedecor, G. & Cochran W. "Métodos estadísticos" C.E.C.S.A. 1979. pp 703
- [31] Taha, H. "Operation Research: an Introduction"m Macmillan. 1976. pp 648
- [32] Wolper D. & Macready W. "No Free Lunch Theorems for Optimixation" Proceedings of the IEEE International Conference on Evolutionary Computation 1997. p 67-82.