

01171
lej.1



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DIVISION DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERIA

Teoría de Algoritmos de Programación no lineal

TESIS DE MAESTRIA
QUE PARA OBTENER EL GRADO DE
MAESTRO EN INVESTIGACION DE OPERACIONES

presenta

FRANCISCO VENEGAS MARTINEZ

TESIS CON
FALLA DE ORIGEN



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 está 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.

INDICE

pag.

INTRODUCCION 1

CAPITULO 1 ELEMENTOS DE LA PROGRAMACION NO LINEAL

1.1 ANALISIS CONVEXO

1.1.1 INTRODUCCION 5

1.1.2 CONJUNTOS CONVEXOS 6

1.1.3 FUNCIONES CONVEXAS Y SU
GENERALIZACION 30

1.2 CONDICIONES DE OPTIMALIDAD

1.2.1 INTRODUCCION 54

1.2.2 CONDICIONES DE FRITZ JOHN
Y KUHN-TUCKER 55

CAPITULO 2 FUNDAMENTOS DE LA TEORIA DE ALGORITMOS DE PROGRAMACION NO LINEAL

2.1 INTRODUCCION 70

2.2 MAPEOS ALGORITMICOS Y FUNCIONES DE
DESCENSO 70

	pag.
2.3 MAPEOS CERRADOS	77
CAPITULO 3	
CONVERGENCIA DE MAPEOS ALGORITMICOS	
3.1 INTRODUCCION	84
3.2 EL TEOREMA FUNDAMENTAL DE CONVERGENCIA ALGORITMICA	89
3.3 COMPOSICION DE MAPEOS	90
3.4 RAPIDEZ DE CONVERGENCIA	95
CAPITULO 4	
EL ALGORITMO DE PAVIANI	
4.1 INTRODUCCION	100
4.2 METODO DE BUSQUEDA POLIEDRICA	100
4.3 DESCRIPCION DEL ALGORITMO DE PAVIANI	108
4.4 ESTRUCTURA DEL PROGRAMA ALGPAY	115
4.5 ENTRADA DE DATOS DEL PROGRAMA ALGPAY	117
4.6 USO DEL PROGRAMA ALGPAY	119
4.7 PROGRAMA FORTRAN ALGPAY	125
BIBLIOGRAFIA	139

Estuve en una escuela de Matemáticas, en donde el profesor enseñaba a sus alumnos con un método apenas imaginable para nosotros en Europa. La proposición y demostración eran escritas sobre una oblea delgada con tintura cefálica, la oblea tenía que ser tomada en ayunas. Durante los tres días siguientes nada más se podía ingerir pan y agua. Cuando la oblea era digerida la tintura ascendía al cerebro, llevando la proposición y demostración junto con ella.

Jonathan Swift en

Viajes de Gulliver.

INTRODUCCION

En 1970 Zangwill y Polak, utilizando las nociones de semicontinuidad y de adherencia presentaron por primera vez un conjunto de resultados generales sobre condiciones suficientes para la convergencia de una extensa clase de algoritmos descendentes de programación no lineal, motivando con esto, el desarrollo de una teoría general de convergencia algorítmica. Poco después en 1975, Huard aplicando el concepto de mapeo cerrado estudió la convergencia de algunos algoritmos de programación no lineal, dando pauta a que Meyer y Komlósi, recientemente, combinando la noción de compacidad con la de mapeo cerrado, obtuvieran resultados todavía más generales que los de Zangwill y Polak.

Pocos han sido los resultados relevantes que se han obtenido sobre la convergencia de algoritmos de programación no lineal y muchos los resultados aparentemente aislados o sin conexión. El principal objetivo que lleva consigo este trabajo, consiste en fundamentar y desarrollar una teoría general sobre la convergencia de algoritmos de programación no lineal que conjunte y conecte dichos resultados, y que trate fundamentalmente los dos aspectos siguientes:

(i). Convergencia global.

Dado cualquier punto inicial, se cuestiona si el algoritmo converge.

(ii). Rapidez de convergencia.

Dados dos algoritmos convergentes. ¿Cuál de los dos converge más rápido ?.

Un hecho significativo para el desarrollo adecuado de una teoría general de algoritmos de programación no lineal, consistirá en asociar a cada algoritmo un mapeo de punto a conjunto; Un algoritmo será considerado como un procedimiento iterativo que genera una sucesión de puntos de acuerdo a un conjunto de reglas preestablecidas y a la información cedida por puntos previamente obtenidos.

Muchos de los algoritmos empleados en programación no lineal, por su complejidad, pueden ser considerados como la composición de dos o más mapeos de punto a conjunto, siendo absolutamente necesario, incluir dentro de la teoría el estudio de mapeos composición y de su convergencia.

En el transcurso del capítulo 1, se presentan los conceptos y resultados más relevantes del análisis convexo, casi indispensables, en la obtención de condiciones de optimalidad en programación no lineal. A través del capítulo 2, se presentan los fundamentos necesarios para el desarrollo adecuado de una teoría general que trate sobre la convergencia de algoritmos de programación no lineal. En el capítulo 3, se estudia la convergencia algorítmica y la rapidez con que esta se efectúa. Finalmente, en el capítulo 4, es estudiado un algoritmo de programación no lineal desarrollado, recientemente, por D. Pavianni.

CAPITULO

1

ELEMENTOS DE LA PROGRAMACION
NO LINEAL

CAPITULO 1. ELEMENTOS DE LA PROGRAMACION NO LINEAL

1.1 ANALISIS CONVEXO

1.1.1 INTRODUCCION

Uno de los tópicos de la matemática, ya sea pura o aplicada, que ha cobrado un creciente interés en las últimas décadas, es el análisis convexo. La ingeniería y la investigación de operaciones han dirigido también su atención en forma ascendente a la aplicación del análisis convexo en la teoría de optimización.

El primer estudio formal sobre el tema de convexidad, apareció en 1911, con la obra de Herman Minkowski, titulada: " Teoría de cuerpos convexos ". El segundo trabajo importante sobre convexidad fue realizado por Werner Fenchel y fue publicado en 1934 bajo el título: " Sobre funciones convexas conjugadas ". Casi desde entonces, fue reconocido el importante papel que desempeñan las funciones convexas en la teoría de optimización debido a sus particulares propiedades. Sin embargo, es importante hacer notar aquí, que en la práctica, es imposible esperar que todas las funciones sean convexas; pero lo que si es posible, es investigar sobre la generalización de dichas funciones, relajando los supuestos, de tal manera, que las propiedades deseables para efectos de optimalidad se preserven.

En el estudio de la Programación no lineal, el análisis convexo participa esencialmente en la obtención de condiciones adecuadas de optimalidad bajo hipótesis relajadas de convexidad.

1.1.2 CONJUNTOS CONVEXOS

En esta sección, se presentan los conceptos y resultados más relevantes del análisis convexo, los cuales comprenden: En primer término. El teorema de Carathéodory sobre la compacidad de la envolvente convexa de conjuntos compactos, y la noción de envolvente convexo-balanceada útil en la construcción de conjuntos convexos con cierta simetría con respecto al origen. Después se estudian las propiedades topológicas de los conjuntos convexos y de las envolventes convexa y convexo-balanceada. A continuación se presentan los teoremas de hiperplanos de soporte en la frontera de conjuntos convexos y los teoremas de separación entre conjuntos convexos y ajenos por hiperplanos, como consecuencia directa de estos últimos resultados se establecen los teoremas de alternativa de Farkas y Gordan.

En todo lo que sigue se hará referencia al espacio euclidiano n -dimensional \mathbb{R}^n y al producto escalar $\langle x, y \rangle = x^t y$, en donde x , y son vectores columna de \mathbb{R}^n . Este producto escalar define en \mathbb{R}^n la norma $\|x\|^2 = \langle x, x \rangle$ y esta norma a su vez define la métrica $d(x, y) = \|x - y\|$, la cual induce una topología τ , en la que los conjuntos abiertos son de la forma $\{y \in \mathbb{R}^n; \|x - y\| < \epsilon\}$ con $\epsilon > 0$. La siguiente definición relacionada con τ será de utilidad en el desarrollo de esta sección.

Definición. 1.1.2.1 Sea A un subconjunto no vacío de \mathbb{R}^n , entonces:

(i). La cerradura de A , denotada por \bar{A} , es la intersección de todos los conjuntos cerrados que contienen a A . Un punto $x \in \mathbb{R}^n$ pertenece a \bar{A} si, y sólo si $\{y \in \mathbb{R}^n; \|x-y\| < \epsilon\} \cap A \neq \emptyset$ para todo $\epsilon > 0$.

(ii). El interior de A , denotado por $\text{Int}(A)$, es la unión de todos los conjuntos abiertos contenidos en A . Un punto $x \in \mathbb{R}^n$ pertenece a $\text{Int}(A)$ si, y sólo si, existe $\epsilon > 0$, tal que $\{y \in \mathbb{R}^n; \|x-y\| < \epsilon\} \subset A$.

(iii). La frontera de A , denotada por ∂A , es el conjunto $\bar{A} \setminus \text{Int}(A)$. Un punto $x \in \mathbb{R}^n$ pertenece a ∂A si, y sólo si $\{y \in \mathbb{R}^n; \|x-y\| < \epsilon\} \cap A \neq \emptyset$ y $\{y \in \mathbb{R}^n; \|x-y\| < \epsilon\} \cap A^c \neq \emptyset$.

(iv). Se dice que un conjunto $A \subset \mathbb{R}^n$ es compacto, si toda cubierta abierta de A contiene una subcubierta finita, es decir, si $\{U_\alpha\}_{\alpha \in I}$ es cualquier familia de conjuntos abiertos, tal que $A \subset \bigcup_{\alpha \in I} U_\alpha$, entonces existe una subfamilia finita $\{U_{\alpha_1}, U_{\alpha_2}, \dots, U_{\alpha_k}\}$ tal que $A \subset \bigcup_{i=1}^k U_{\alpha_i}$.

Los conceptos fundamentales sobre los cuales descansa la esencia toda del pensamiento matemático es el concepto de límite de una sucesión infinita. En el estudio de la matemática, este concepto aparece casi invariablemente, ya sea implícita o explícitamente.

Definición. 1.1.2.2

(i). Una sucesión $\{x_k\}_{k \in \mathbb{N}}$ de elementos de \mathbb{R}^n , se dice que converge a un elemento $x^* \in \mathbb{R}^n$, si dado $\epsilon > 0$, existe $N \in \mathbb{N}$, tal que para toda $k \in \mathbb{N}$ con $k > N$, se tiene que $\|x_k - x^*\| < \epsilon$. Se denota que la sucesión $\{x_k\}_{k \in \mathbb{N}}$ converge a $x^* \in \mathbb{R}^n$ por $\lim_{k \rightarrow \infty} x_k = x^*$. Frecuentemente, este hecho se denota como: $\|x_k - x^*\| \rightarrow 0$ cuando $k \rightarrow \infty$, o bien, $x_k \xrightarrow{\|\cdot\|} x^*$ cuando $k \rightarrow \infty$.

(ii). Una sucesión $\{x_k\}_{k \in \mathbb{N}}$ de elementos de \mathbb{R}^n , es llamada de Cauchy, si dado $\epsilon > 0$, existe $N \in \mathbb{N}$, tal que para todo $k, m \in \mathbb{N}$ con $k, m > N$, se tiene que $\|x_k - x_m\| < \epsilon$.

Cabe mencionar aquí, que en \mathbb{R}^n toda sucesión de Cauchy converge.

Proposición. 1.1.2.3 Sea A un subconjunto no vacío de \mathbb{R}^n . Entonces las siguientes condiciones son equivalentes:

- (i). A es compacto
- (ii). A es cerrado y acotado
- (iii). Toda sucesión de elementos de A contiene una subsucesión convergente

La prueba de la proposición anterior se omite, y puede ser encontrada en la referencia

Definición. 1.1.2.4 Sean $A, B \subset \mathbb{R}^n$ y $\lambda \in \mathbb{R}$, entonces

- (i). $A+B = \{x+y ; x \in A, y \in B\}$
- (ii). $\lambda A = \{\lambda x ; \lambda \in \mathbb{R}, x \in A\}$

Frecuentemente se utiliza la siguiente notación: $\{a\}+B = a+B$ y $\{\lambda\}A = \lambda A$.

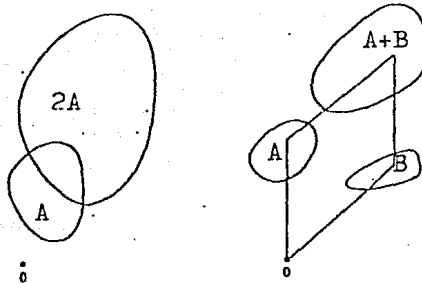


Figura 1. Suma y producto por escalares entre conjuntos.

Definición. 1.1.2.5 Sea $C \subset \mathbb{R}^n$, se dice que el conjunto C es convexo, si $\lambda C + (1-\lambda)C \subset C$ para todo $\lambda \in [0,1]$.

Equivalentemente, si $x, y \in C$, entonces $\lambda x + (1-\lambda)y \in C$ para todo $\lambda \in [0,1]$.

Geométricamente, lo que la definición anterior dice, es que un conjunto C es convexo, si dados dos puntos cualesquiera del conjunto, los puntos del segmento de recta que los une pertenecen también al conjunto.

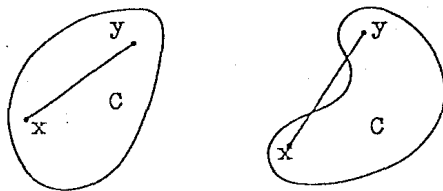


Figura 2. Convexidad

Aunque la definición de conjunto convexo se expresa en términos de pares de puntos. Es posible mostrar en forma sencilla, que un conjunto es convexo si, y sólo si cada combinación lineal convexa de un número finito de puntos del conjunto es de nuevo un punto del conjunto.

Observe también que la definición de conjunto convexo, tiene sentido en cualquier espacio vectorial.

Las propiedades mencionadas en la siguiente proposición son sencillas de probar o casi triviales, por lo que se omite su prueba. Dichas propiedades son útiles, cuando es inconveniente verificar convexidad directamente a partir de la definición.

Proposición. 1.1.2.6

- (i). La intersección de cualquier familia de conjuntos convexos es de nuevo un conjunto convexo.
- (ii). Una combinación lineal finita de conjuntos convexos es de nuevo un conjunto convexo.

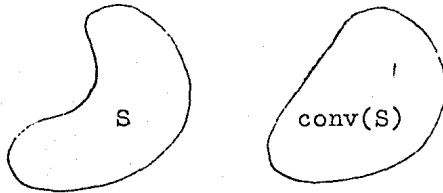
Proposición. 1.1.2.7 Sea $C \subset \mathbb{R}^n$ un conjunto convexo y $\alpha, \beta > 0$, entonces $(\alpha + \beta)C = \alpha C + \beta C$.

Prueba:

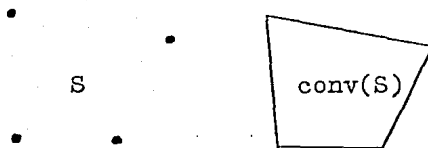
Si $\alpha=0$ ó $\beta=0$ no hay nada que probar. Suponga que $\alpha, \beta > 0$, puesto que C es un conjunto convexo, entonces $\frac{\alpha}{\alpha+\beta}C + \frac{\beta}{\alpha+\beta}C \subset C$, de donde $\alpha C + \beta C \subset (\alpha+\beta)C$, la contención contraria se satisface para cualquier conjunto C y para cualesquier escalares α y β .

Dado un subconjunto S de \mathbb{R}^n se sabe que existen subconjuntos convexos de \mathbb{R}^n que lo contienen, por ejemplo, el mismo \mathbb{R}^n , y se sabe además por la proposición 1.1.2.6, que la intersección de conjuntos convexos es de nuevo un conjunto convexo, en consecuencia, se sigue que existe un mínimo conjunto convexo que contiene a S . Este hecho se precisa a continuación:

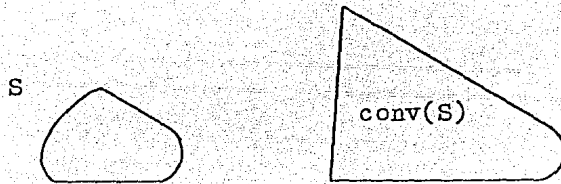
Definición. 1.1.2.8 Sea $S \subset \mathbb{R}^n$. La envolvente convexa de S denota da por $\text{conv}(S)$, es la intersección de todos los conjuntos convexos que contienen a S .



(a)



(b)



(c)

Figura 3. Los incisos (a), (b) y (c) son ejemplos de envolventes convexas.

Proposición. 1.1.2.9 Sea $S \subset \mathbb{R}^n$, entonces

(i). $\text{conv}(S)$ es el mínimo conjunto convexo que contiene a S

(ii). $\text{conv}(S) = \left\{ \sum_{i=1}^m \alpha_i x_i \in \mathbb{R}^n; \alpha_i \in [0, 1], \sum_{i=1}^m \alpha_i = 1, x_i \in S, m \in \mathbb{N} \right\}$

Prueba:

(i) se sigue de la proposición 1.1.2.6;

(ii). Sea $T = \left\{ \sum_{i=1}^m \alpha_i x_i \in \mathbb{R}^n; \alpha_i \in [0, 1], \sum_{i=1}^m \alpha_i = 1, x_i \in S, m \in \mathbb{N} \right\}$. Claramente $S \subset T$, se verá que T es un conjunto convexo, entonces por (i) $\text{conv}(S) \subset T$. Sean $y_1, y_2 \in T$ y $\lambda \in [0, 1]$, entonces $y_1 = \sum_{i=1}^m \alpha_i x_i$ y $y_2 = \sum_{i=1}^m \alpha'_i x_i$, en donde algunos α_i ó α'_i son cero, de tal manera que y_1 y y_2 se expresen como combinaciones convexas de los mismos elementos (si es necesario). Entonces $\lambda y_1 + (1-\lambda)y_2 = \sum_{i=1}^m (\lambda \alpha_i + (1-\lambda)\alpha'_i) x_i \in T$, ya que $\sum_{i=1}^m \lambda \alpha_i + (1-\lambda)\alpha'_i = 1$ y $\lambda \alpha_i + (1-\lambda)\alpha'_i \geq 0$. Por lo tanto T es convexo y en consecuencia $\text{conv}(S) \subset T$.

Finalmente, se probará por inducción sobre m que $T \subset \text{conv}(S)$.

Para $m=1$, se sigue de que $S \subset \text{conv}(S)$. Suponga que para $k \leq m-1$, $\sum_{i=1}^k \alpha_i x_i \in \text{conv}(S)$ con $\alpha_i \geq 0$, $\sum_{i=1}^k \alpha_i = 1$ y $x_i \in S$. Sea $y = \sum_{i=1}^m \alpha'_i x_i$ con $\alpha'_i \geq 0$, $\sum_{i=1}^m \alpha'_i = 1$ y $x_i \in S$, si algún $\alpha'_1 = 0$, entonces por hipótesis de induc

ción $ye \text{conv}(S)$, de donde es posible suponer que $\alpha_i' > 0$ para toda i . Sea $\alpha = \sum_{i=1}^{m-1} \alpha_i' > 0$, entonces $\sum_{i=1}^{m-1} \frac{\alpha_i'}{\alpha} = 1$ y $1 - \alpha = \alpha_m'$. Por hipótesis de inducción $\sum_{i=1}^{m-1} \frac{\alpha_i'}{\alpha} x_i \in \text{conv}(S)$ y $x_m \in \text{conv}(S)$ y puesto que $\text{conv}(S)$ es un conjunto convexo, $y = \alpha \sum_{i=1}^{m-1} \frac{\alpha_i'}{\alpha} x_i + (1 - \alpha)x_m \in \text{conv}(S)$. Por lo tanto $T \subset \text{conv}(S)$.

De acuerdo a la proposición anterior, un punto en la envolvente convexa de un conjunto puede ser representado como combinación lineal convexa de un número finito de puntos. El siguiente teorema debido a Constantin Carathéodory muestra que cualquier punto de la envolvente convexa de un conjunto SCR^n puede ser representado como combinación lineal convexa de, a lo más $n+1$ puntos de S .

Teorema 1.1.2.10 (Carathéodory) Sea SCR^n , entonces para todo $y \in \text{conv}(S)$, existen $\{x_1, x_2, \dots, x_k\} \subset S$ y $\{\lambda_1, \lambda_2, \dots, \lambda_k\} \subset \mathbb{R}$, tales que

a) $k \leq n+1$

b) $y = \sum_{i=1}^k \lambda_i x_i$

c) $\lambda_i \geq 0$ y $\sum_{i=1}^k \lambda_i = 1$

Prueba:

Sea $y \in \text{conv}(S)$, entonces $y = \sum_{i=1}^k \lambda_i x_i$, donde $\lambda_i \geq 0$, $\sum_{i=1}^k \lambda_i = 1$ y $x_i \in S$. Si $k \leq n+1$, no hay nada que probar. Suponga que $k > n+1$, entonces $x_2 - x_1, x_3 - x_1, \dots, x_k - x_1$ son necesariamente vectores linealmente dependientes en \mathbb{R}^n , por lo tanto existen escalares $\mu_2, \mu_3, \dots, \mu_k$ no todos cero, tales que $\sum_{i=2}^k \mu_i (x_i - x_1) = 0$. Defina $\mu_1 = -\sum_{i=2}^k \mu_i$, entonces $\sum_{i=1}^k \mu_i x_i = \sum_{i=2}^k \mu_i x_i - \mu_1 x_1 = \sum_{i=2}^k \mu_i x_i - \sum_{i=2}^k \mu_i x_1 = \sum_{i=2}^k \mu_i (x_i - x_1) = 0$ y $\sum_{i=1}^k \mu_i = 0$ con no todos los μ_i nulos, además por lo menos un $\mu_i > 0$. Entonces para toda $\alpha \in \mathbb{R}$ se cumple que $y = \sum_{i=1}^k \lambda_i x_i - 0 = \sum_{i=1}^k \lambda_i x_i - \alpha \sum_{i=1}^k \mu_i x_i = \sum_{i=1}^k (\lambda_i - \alpha \mu_i) x_i$. Escoja α como sigue: $\alpha = \min_{\substack{1 \leq i \leq k \\ \mu_i > 0}} \left\{ \frac{\lambda_i}{\mu_i}; \mu_i > 0 \right\} = \frac{\lambda_j}{\mu_j} > 0$. Ahora bien, si $\mu_i \leq 0$, entonces $\alpha \mu_i \leq 0 < \lambda_i$, así $\lambda_i - \alpha \mu_i > 0$ y si $\mu_i > 0$, entonces $\frac{\lambda_i}{\mu_i} \geq \frac{\lambda_j}{\mu_j} = \alpha$ y por lo tanto $\lambda_i - \alpha \mu_i \geq 0$. En conclusión $\lambda_i - \alpha \mu_i \geq 0$ para toda i y en particular $\lambda_j - \alpha \mu_j = 0$ debido a la definición de α . Por lo tanto $y = \sum_{i=1}^k (\lambda_i - \alpha \mu_i) x_i$ con $\lambda_i - \alpha \mu_i \geq 0$ para

para toda i , $\sum_{i=1}^k (\lambda_i - \alpha \mu_i) = 1$, $\{x_1, x_2, \dots, x_k\} \subset S$ y $\lambda_j - \alpha \mu_j = 0$. En otras palabras $y \in \text{conv}(S)$ se puede expresar como combinación lineal convexa de $k-1$ elementos de S . Si $k-1 = n+1$ la prueba termina. En caso contrario, se repite el procedimiento anterior hasta representar a $y \in \text{conv}(S)$ como combinación lineal convexa de $n+1$ elementos.

La envolvente convexa de un subconjunto S de \mathbb{R}^n posee muchas propiedades geométricas y topológicas notables, las cuales serán expuestas a continuación:

Corolario. 1.1.2.11 Sea $S \subset \mathbb{R}^n$, si S es un conjunto compacto, entonces $\text{conv}(S)$ es también compacto.

Prueba:

Observe primero que la función $F: \mathbb{R}^{n+1} \times (\mathbb{R}^n)^{n+1} \rightarrow \mathbb{R}^n$ definida mediante $F(\lambda_1, \lambda_2, \dots, \lambda_{n+1}; x_1, x_2, \dots, x_{n+1}) = \sum_{i=1}^{n+1} \lambda_i x_i$ es una función continua y que el conjunto $A = \{(\lambda_1, \lambda_2, \dots, \lambda_{n+1}); \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1\}$ es compacto. Sea $Q = \underbrace{S \times S \times \dots \times S}_{n+1}$, entonces AXQ es compacto puesto que el producto cartesiano de conjuntos compactos es de nuevo un conjunto compacto, así $F(AXQ)$ es un conjunto compacto de \mathbb{R}^n , ya que, éste es la imagen continua de un compacto. Por el teorema de Carathéodory se tiene que $\text{conv}(S) \subset F(AXQ)$. Por otra parte, si $y \in F(AXQ)$, entonces $y = \sum_{i=1}^{n+1} \lambda_i x_i$, $\lambda_i \geq 0$, y $\sum_{i=1}^{n+1} \lambda_i = 1$, lo cual implica que $y \in \text{conv}(S)$, de donde $F(AXQ) \subset \text{conv}(S)$ y por lo tanto $\text{conv}(S)$ es compacto.

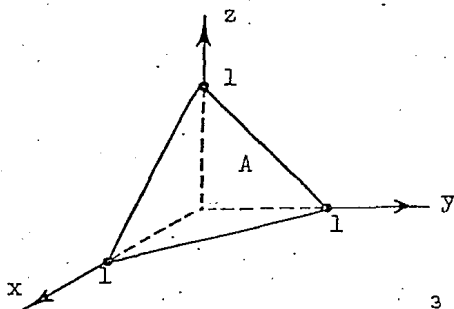


Figura 4. $A = \{(\lambda_1, \lambda_2, \lambda_3); \lambda_i \geq 0, \sum_{i=1}^3 \lambda_i = 1\}$

Proposición. 1.1.2.12 Sean SCR^n y $a \in \mathbb{R}$, entonces $\text{conv}(S+a) = \text{conv}(S)+a$.

Prueba:

Puesto que $S \subset \text{conv}(S)$, entonces $S+a \subset \text{conv}(S)+a$, por la proposición 1.1.2.6, se sigue que $\text{conv}(S)+a$ es un conjunto convexo, entonces por la proposición 1.1.2.9 $\text{conv}(S+a) \subset \text{conv}(S)+a$.

Análogamente $S = (S+a) - a \subset \text{conv}(S+a) - a$, de donde $\text{conv}(S) \subset \text{conv}(S+a) - a$, equivalentemente $\text{conv}(S)+a \subset \text{conv}(S+a)$.

Proposición. 1.1.2.13 Sean $S, T \subset \mathbb{R}^n$, entonces $\text{conv}(S \cup T) =$

$$\bigcup_{\lambda \in [0,1]} (\lambda S + (1-\lambda)T).$$

Prueba:

Denote por $H = \bigcup_{\lambda \in [0,1]} (\lambda S + (1-\lambda)T)$, claramente $S \cup T \subset H$, se verá que el conjunto H es convexo, de donde por la proposición 1.1.2.9 $\text{conv}(S \cup T) \subset H$. Sean $x, y \in H$ y $\alpha \in [0,1]$, entonces existen $\lambda_0, \lambda_1 \in [0,1]$ tales que $x \in \lambda_0 S + (1-\lambda_0)T$ y $y \in \lambda_1 S + (1-\lambda_1)T$, entonces $\alpha x + (1-\alpha)y \in \alpha \lambda_0 S + (1-\alpha)\lambda_1 S + \alpha(1-\lambda_0)T + (1-\alpha)(1-\lambda_1)T = (\alpha \lambda_0 + (1-\alpha)\lambda_1)S + (\alpha(1-\lambda_0) + (1-\alpha)(1-\lambda_1))T$, ésto último se sigue de que S y T son conjuntos convexos (Ver proposición 1.1.2.7). Ahora bien, si $\beta = \alpha \lambda_0 + (1-\alpha)\lambda_1$, entonces $1-\beta = \alpha(1-\lambda_0) + (1-\alpha)(1-\lambda_1)$ con $\beta \in [0,1]$, de donde $\alpha x + (1-\alpha)y \in H$. Por lo tanto, $\text{conv}(S \cup T) \subset H$. Por otro lado, si $x \in H$, entonces existe $\lambda_0 \in [0,1]$, tal que $x = \lambda_0 s + (1-\lambda_0)t$ con $s, t \in S \cup T$, pero $S \cup T \subset \text{conv}(S \cup T)$ y como $\text{conv}(S \cup T)$ es convexo, se sigue que $x \in \text{conv}(S \cup T)$, así $H \subset \text{conv}(S \cup T)$.

Una de las propiedades topológicas más importante que tiene la envolvente convexa de un conjunto S , es que si S es un conjunto abierto, entonces $\text{conv}(S)$ también lo es, como lo muestra el siguiente teorema:

Teorema. 1.1.2.14 Sea SCR^n un conjunto abierto y no vacío, entonces $\text{conv}(S)$ es abierto.

Prueba:

Sea $x \in \text{conv}(S)$, entonces $x = \sum_{i=1}^m \lambda_i x_i$ con $\lambda_i \geq 0$, $\sum_{i=1}^m \lambda_i = 1$, y $x_i \in S$, $m \in \mathbb{N}$.

Si S es abierto, entonces existen $\epsilon_1, \epsilon_2, \dots, \epsilon_m > 0$, tales que $B(x_i; \epsilon_i) = \{y \in \mathbb{R}^n; \|x_i - y\| < \epsilon_i\} \subset S$ para todo i . Sea $\epsilon = \min_{1 \leq i \leq m} \{\epsilon_i\}$, entonces $B(x_i; \epsilon) \subset S$ para todo i , y si $y \in \mathbb{R}^n$ es tal que $\|y\| < \epsilon$, entonces $\|\sum_{i=1}^m \lambda_i x_i - \sum_{i=1}^m \lambda_i (x_i + y)\| = \|y\| < \epsilon$ con $\sum_{i=1}^m \lambda_i (x_i + y) \in \text{conv}(S)$, ya que, $\|(x_i + y) - x_i\| = \|y\| < \epsilon$ para todo i , de donde $x_i + y \in B(x_i; \epsilon) \subset S$ para todo i .

Observe que si SCR^n es cerrado, entonces $\text{conv}(S)$ no es necesariamente cerrado. En efecto, considere en \mathbb{R}^2 el conjunto que consiste de los puntos $(-1, 0), (1, 0)$ y el eje de las ordenadas, la envolvente convexa de este conjunto se muestra en la siguiente figura:

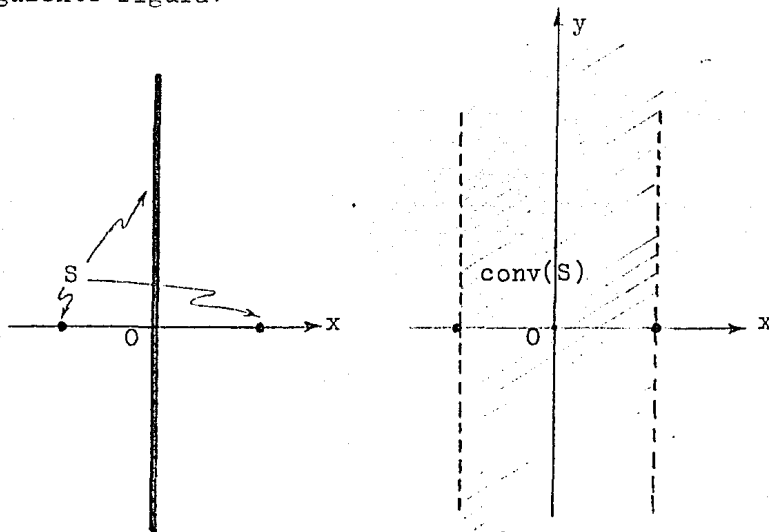


Figura 5. Ejemplo de envolvente convexa

A continuación se discute sobre algunas otras propiedades topológicas de los conjuntos convexos.

Proposición. 1.1.2.15 Sea $C \subset \mathbb{R}^n$ un conjunto convexo, tal que $\text{Int}(C) \neq \emptyset$ y sea $y \in \bar{C}$ y $z \in \text{Int}(C)$. Entonces $y = \lambda x + (1 - \lambda)z \in \text{Int}(C)$ para todo $\lambda \in (0, 1)$.

Prueba:

Si $z \in \text{Int}(C)$, entonces existe $\epsilon > 0$, tal que $\{w \in \mathbb{R}^n; \|w-z\| < \epsilon\} \subset C$.

Se probará que $\{w \in \mathbb{R}^n; \|w-y\| < (1-\lambda)\epsilon\} \subset C$. Sea $w_0 \in \mathbb{R}^n$, tal que $\|w_0-y\| < (1-\lambda)\epsilon$. Puesto que $x \in \bar{C}$, se sigue que $\{w \in \mathbb{R}^n; \|w-x\| < ((1-\lambda)\epsilon - \|w_0-y\|)/\lambda\} \cap C \neq \emptyset$, de donde existe $w_1 \in \mathbb{R}^n$, tal que $\|w_1-x\| < ((1-\lambda)\epsilon - \|w_0-y\|)/\lambda$. Sea $w_2 = \frac{w_0 - \lambda w_1}{1-\lambda}$, entonces $\|w_2-z\| =$

$\left\| \frac{w_0 - \lambda w_1}{1-\lambda} - z \right\| = \left\| \frac{w_0 - \lambda w_1 - y + \lambda x}{1-\lambda} \right\| = \frac{1}{1-\lambda} \|(w_0 - y) + \lambda(x - w_1)\| \leq \frac{1}{1-\lambda} (\|w_0 - y\| + \lambda \|x - w_1\|) < \epsilon$, de donde $w_2 \in C$, pero $w_0 = (1-\lambda)w_2 + \lambda w_1$ y como $w_1 \in C$, esto implica que $w_0 \in C$. Por lo tanto $\{w \in \mathbb{R}^n; \|w-y\| < (1-\lambda)\epsilon\} \subset C$ para todo $\lambda \in (0,1)$ con $y = \lambda x + (1-\lambda)z$.

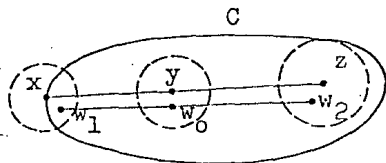


Figura 6. Segmento de línea que une un punto de $\text{Int}(C)$ con uno de ∂C .

Corolario. 1.1.2.16 Sea $C \subset \mathbb{R}^n$ un conjunto convexo, tal que $\text{Int}(C) \neq \emptyset$. Entonces

- (i). $\text{Int}(C)$ es un conjunto convexo
- (ii). \bar{C} es convexo
- (iii). $\overline{\text{Int}(C)} = \bar{C}$
- (iv). $\text{Int}(\bar{C}) = \text{Int}(C)$

Prueba:

(i). Sean $x, z \in \text{Int}(C)$ y como $\text{Int}(C) \subset \bar{C}$, en particular $z \in \bar{C}$, de donde $\lambda x + (1-\lambda)z \in \text{Int}(C)$ para toda $\lambda \in (0,1)$.

(ii). Sean $x, y \in \bar{C}$ y sea $z \in \text{Int}(C)$, entonces $\lambda x + (1-\lambda)z \in \text{Int}(C)$ para toda $\lambda \in (0,1)$, sea $\mu \in (0,1)$ fijo, entonces $\mu y + (1-\mu)(\lambda x + (1-\lambda)z) \in \text{Int}(C) \subset C$ para toda $\lambda \in (0,1)$. Si ahora se toma el límite cuando $\lambda \rightarrow 1$ se tendrá entonces que $\mu y + (1-\mu)x \in \bar{C}$.

(iii). Claramente $\overline{\text{Int}(C)} \subset \bar{C}$. Sea $x \in \bar{C}$ y $z \in \text{Int}(C)$, entonces

$\lambda x + (1-\lambda)z \in \text{Int}(C)$ para toda $\lambda \in (0,1)$. Si $\lambda \rightarrow 1$, entonces $x \in \overline{\text{Int}(C)}$. Por lo tanto $\bar{0} \subset \overline{\text{Int}(C)}$.

(iv). Obviamente $\text{Int}(C) \subset \text{Int}(\bar{C})$. Sea $x \in \text{Int}(\bar{C})$, entonces existe $\epsilon > 0$, tal que $\|y-x\| < \epsilon$ implica que $y \in \bar{C}$. Sea $z \in \text{Int}(C)$, $z \neq x$ y sea $y = (1+\alpha)x + \alpha z$, donde $\alpha = \frac{\epsilon}{2\|x-z\|}$, así $\|y-x\| = \|(1+\alpha)x - x + \alpha z\| = \alpha\|x-z\| = \frac{\epsilon}{2} < \epsilon$ implica que $y \in \bar{C}$, pero $x = \lambda y + (1-\lambda)z$ donde $\lambda = \frac{1}{1+\alpha} \in (0,1)$, en consecuencia $x \in \text{Int}(C)$. Por lo tanto $\text{Int}(\bar{C}) \subset \text{Int}(C)$.

A continuación se presenta la noción de envolvente convexo-balanceada, la cual es de utilidad en la construcción de conjuntos convexos con cierta simetría con respecto a origen.

Definición. 1.1.2.17 Sea $S \subset \mathbb{R}^n$, se dice que S es un conjunto balanceado si $\lambda x \in S$ siempre que $x \in S$ y $\lambda \in \mathbb{R}$ con $|\lambda| \leq 1$. Equivalentemente, si $\lambda S \subset S$ siempre que $|\lambda| \leq 1$.

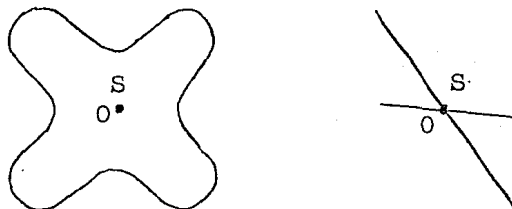


Figura 7. Ejemplos de conjuntos balanceados

Proposición. 1.1.2.18 La intersección y la unión de cualquier familia de conjuntos balanceados son de nuevo conjuntos balanceados.

La prueba de esta proposición es trivial.

De igual manera que tiene sentido hablar de la envolvente convexa de un conjunto, también tiene sentido hablar de la envolvente balanceada, ya que dado un subconjunto S de \mathbb{R}^n , es claro que existen conjuntos balanceados que lo contienen, por ejemplo, el mismo \mathbb{R}^n . Por la proposición anterior se sabe que la

intersección de conjuntos balanceados es de nuevo un conjunto balanceado. En consecuencia existe un mínimo conjunto balanceado que contiene a S .

Definición. 1.1.2.19 Sea SCR^n , la envolvente balanceada de S , denotada por $bal(S)$, es la intersección de todos los conjuntos balanceados que contienen a S .

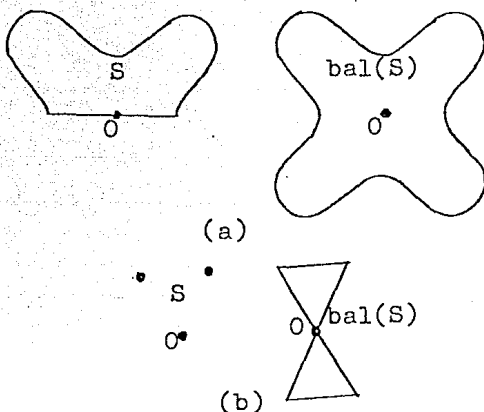


Figura 8. Ejemplos de envolventes balanceadas.

Proposición. 1.1.2.20 Sea SCR^n . Entonces

(i). $bal(S)$ es el mínimo conjunto balanceado que contiene a S

(ii). $bal(S) = \{ \lambda x \in \mathbb{R}^n; \lambda \in \mathbb{R} \text{ con } |\lambda| \leq 1 \text{ y } x \in S \}$

(iii). $bal(S) = \bigcup_{|\lambda| \leq 1} \lambda S$

Prueba:

(i) se sigue de la proposición 1.1.2.18.

(ii). Denote $T = \{ \lambda x \in \mathbb{R}^n; \lambda \in \mathbb{R} \text{ con } |\lambda| \leq 1 \text{ y } x \in S \}$ y por $\Lambda = \{ \lambda \in \mathbb{R}; |\lambda| \leq 1 \}$, entonces $T = \Lambda S$, de (i) se sigue que $S \subset bal(S)$ y dado que $bal(S)$ es balanceado, se tiene $T = \Lambda S \subset \Lambda bal(S) \subset bal(S)$.

Por otra parte, claramente $S \subset T$, se verá que T es balanceado y en consecuencia se tendrá que $bal(S) \subset T$. Sea $y \in T$, entonces existen $\mu \in \Lambda$ y $x \in S$ tales que $y = \mu x$. Sea $\lambda \in \Lambda$, entonces $\lambda y = (\lambda \mu) x \in T$, ya que, $\lambda \mu \in \Lambda$, de donde T es balanceado.

(iii). Denote $T = \bigcup_{|\lambda| \leq 1} \lambda S$. Claramente T es balanceado y $S \subset T$, de donde $\text{bal}(S) \subset T$. Por otra parte, si $y \in T$, entonces existen $\mu \in \mathbb{R}$ con $|\mu| \leq 1$ y $x \in S$ tales que $y = \mu x$, pero $S \subset \text{bal}(S)$, teniéndose así que $y = \mu x \in \text{bal}(S)$. Por lo tanto $T \subset \text{bal}(S)$.

El propósito de haber estudiado el concepto de envolvente balanceada de un conjunto dentro del tema de convexidad se refleja en la siguiente definición:

Definición. 1.1.2.21 Sea $S \subset \mathbb{R}^n$, la envolvente convexo-balanceada de S , denotada por $\text{env}_{c,b}(S)$, se define como $\text{env}_{c,b}(S) = \text{conv}(\text{bal}(S))$.

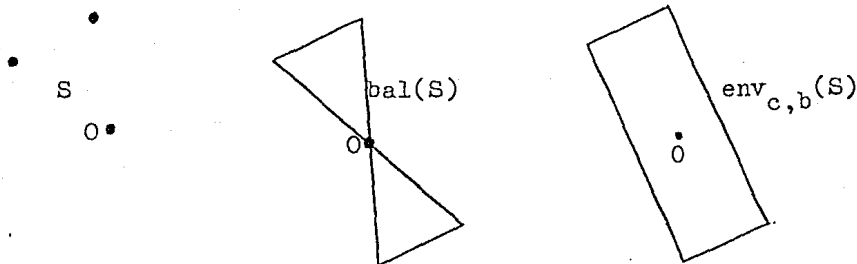


Figura 9. Ejemplo de envolvente convexo-balanceada de S

Se verá en la siguiente proposición que $\text{env}_{c,b}(S)$ es el mínimo conjunto convexo y balanceado que contiene a S . Sin embargo, es importante hacer notar aquí, que $\text{bal}(\text{conv}(S))$ puede no ser un conjunto convexo. En efecto, considere el conjunto integrado por los puntos $(1,0)$ y $(0,-1)$, la envolvente balanceado-convexa se muestra en la siguiente figura:

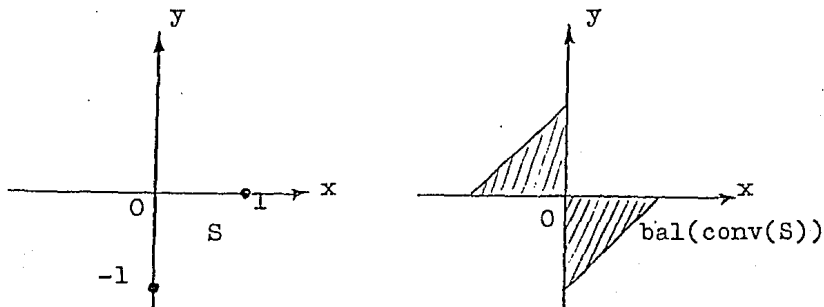


Figura 10. Ejemplo de envolvente balanceado-convexa

Lema. 1.1.2.22 Sea $AC\mathbb{R}^n$ un conjunto balanceado, entonces $\text{conv}(A)$ es balanceado.

Prueba:

Se sabe por la proposición 1.1.2.9 que $\text{conv}(A) = \left\{ \sum_{i=1}^m \lambda_i x_i \in \mathbb{R}^n; \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1, x_i \in A, m \in \mathbb{N} \right\}$. Si $x \in \text{conv}(A)$, entonces $x = \sum_{i=1}^m \lambda_i x_i$ con las propiedades antes mencionadas. Sea $\mu \in \mathbb{R}$, tal que $|\mu| \leq 1$, entonces $\mu x = \sum_{i=1}^m \lambda_i (\mu x_i)$ con $\mu x_i \in A$, ya que el conjunto A es balanceado. Por lo tanto $\mu x \in \text{conv}(A)$ con $|\mu| \leq 1$, siendo así que $\text{conv}(A)$ es balanceado.

Lema. 1.1.2.23 Sean $A, B \subset \mathbb{R}^n$, tales que ACB . Entonces

- (i). $\text{bal}(A) \subset \text{bal}(B)$
- (ii). $\text{conv}(A) \subset \text{conv}(B)$
- (iii). $\text{env}_{c,b}(A) \subset \text{env}_{c,b}(B)$

Prueba:

- (i). Si $y \in \text{bal}(A)$, entonces por la proposición 1.1.2.20 $y = \lambda x$ con $x \in A$ y $\lambda \in \mathbb{R}$, tal que $|\lambda| \leq 1$ y como ACB , se sigue que $y = \lambda x$ con $x \in B$ y $\lambda \in \mathbb{R}$, tal que $|\lambda| \leq 1$. De donde $x \in \text{bal}(B)$.
- (ii). Si $y \in \text{conv}(A)$ en virtud de la proposición 1.1.2.9, se sigue que existe $m \in \mathbb{N}$, tal que $y = \sum_{i=1}^m \lambda_i x_i$ con $\lambda_i \geq 0$, $\sum_{i=1}^m \lambda_i = 1$ y $x_i \in A$ y como ACB , lo anterior también sucede con $x_i \in B$. Así, $\text{conv}(A) \subset \text{conv}(B)$.
- (iii) es consecuencia inmediata de (i) y (ii).

Proposición. 1.1.2.24 Sea $S \subset \mathbb{R}^n$, entonces

- (i). $\text{env}_{c,b}(S)$ es el mínimo conjunto convexo y balanceado que contiene a S .
- (ii). $\text{env}_{c,b}(S)$ es la intersección de todos los conjuntos convexos y balanceados que contienen a S .
- (iii). $\text{env}_{c,b}(S) = \left\{ \sum_{i=1}^m \lambda_i x_i \in \mathbb{R}^n; \lambda_i \in \mathbb{R}, \sum_{i=1}^m |\lambda_i| \leq 1, x_i \in S, m \in \mathbb{N} \right\}$

Prueba:

- (i) Denote $B = \text{env}_{c,b}(S)$. Hay que ver que:
 - (a). B es un conjunto convexo y balanceado
 - (b). $S \subset B$

(c). Si C es un conjunto convexo y balanceado tal que SCC , entonces BCC .

En efecto,

(a). Claramente B es convexo, que sea balanceado se sigue del lema 1.1.2.22.

(b). Es inmediato que $SCbal(S) \subset conv(bal(S)) = B$.

(c). Por el lema 1.1.2.23, si C es convexo y balanceado y SCC , entonces $bal(S) \subset bal(C) = C$ y $conv(bal(S)) \subset conv(C) = C$.

(ii). se sigue de las proposiciones 1.1.2.6 y 1.1.2.18.

(iii). Denote $T = \{ \sum_{i=1}^m \alpha_i x_i \in \mathbb{R}^n; \alpha_i \in \mathbb{R}, \sum_{i=1}^m |\alpha_i| \leq 1, x_i \in S, m \in \mathbb{N} \}$. Hay que ver que:

(d). T es un conjunto convexo

(e). T es un conjunto balanceado

(f). SCT .

En efecto,

(d). Sean $y_1, y_2 \in T$, entonces existe $m \in \mathbb{N}$ tal que $y_1 = \sum_{i=1}^m \alpha_i x_i$ y $y_2 = \sum_{i=1}^m \alpha'_i x_i$, en donde algunos α_i ó α'_i son cero (en caso de que sea necesario), de tal manera que y_1 y y_2 se expresen como combinaciones lineales de los mismos elementos. Si $\beta \in [0, 1]$, entonces $\beta y_1 + (1-\beta)y_2 = \sum_{i=1}^m (\beta\alpha_i + (1-\beta)\alpha'_i)x_i \in T$, ya que,

$$\sum_{i=1}^m |\beta\alpha_i + (1-\beta)\alpha'_i| \leq \beta \sum_{i=1}^m |\alpha_i| + (1-\beta) \sum_{i=1}^m |\alpha'_i| \leq \beta + (1-\beta) = 1.$$

(e). Sea $y \in T$, entonces existe $m \in \mathbb{N}$ tal que $y = \sum_{i=1}^m \lambda_i x_i$. Si $\mu \in \mathbb{R}$ con $|\mu| \leq 1$, entonces $\mu y = \sum_{i=1}^m (\mu\lambda_i)x_i \in T$, ya que $\sum_{i=1}^m |\mu\lambda_i| = |\mu| \sum_{i=1}^m |\lambda_i| \leq 1$.

(f). Obviamente para $m=1$ se tiene que SCT .

De los incisos (d), (e) y (f) se concluye que BCT . Se verá ahora la contención contraria, TcB . Sean $x_i \in S$ y $\alpha_i \in \mathbb{R}$, tales que $\sum_{i=1}^m |\alpha_i| \leq 1$. Ponga $t = \sum_{i=1}^m |\alpha_i|$ y $\beta_i = \frac{\alpha_i t}{|\alpha_i|}$, entonces $|\beta_i| = t \leq 1$, lo cual implica que $\beta_i x_i \in bal(S)$. Sea ahora $\lambda_i = \frac{|\alpha_i|}{t}$, entonces $\sum_{i=1}^m \lambda_i = 1$ y $\lambda_i \in [0, 1]$, de donde $\sum_{i=1}^m \lambda_i \beta_i x_i = \sum_{i=1}^m \frac{|\alpha_i|}{t} \frac{\alpha_i t}{|\alpha_i|} x_i = \sum_{i=1}^m \alpha_i x_i \in B$. Por lo tanto TcB y en consecuencia $T=B$.

La noción de hiperplano de soporte y la propiedad que tienen los conjuntos convexos ajenos de poder ser separados por hiperplanos constituyen una parte importante de la teoría de optimización. Casi todas las condiciones de optimalidad y dualidad las utilizan en algún sentido.

Los resultados que a continuación se derivan están basados en el siguiente hecho geométrico: Dado un conjunto convexo y cerrado CCR^n y un punto $y \notin C$, existe un único punto $x^* \in C$ que minimiza la distancia a y .

Teorema. 1.1.2.25 (Teorema de proyección o Lema de Riesz)
Sea CCR^n un conjunto convexo y cerrado y $y \notin C$. Entonces existe un único punto $x^* \in C$ que minimiza la distancia a y . Algo más, x^* minimiza la distancia a y si, y sólo si $\langle x - x^*, x^* - y \rangle \geq 0$ para toda $x \in C$.

Prueba:

Sea $\delta = \inf \{ \|y - x\|; x \in C \} > 0$ y sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión de elementos de C , tal que $\|y - x_k\| \rightarrow \delta$ cuando $k \rightarrow \infty$. Se verá que $\lim_{k \rightarrow \infty} x_k = x^* \in C$ probando que $\{x_k\}_{k \in \mathbb{N}}$ es una sucesión de Cauchy. En efecto, utilizando la ley del paralelogramo, se tiene que

$$\begin{aligned} \|x_k - x_m\|^2 &= \|(x_k - y) - (x_m - y)\|^2 = 2\|x_k - y\|^2 + 2\|x_m - y\|^2 - \|x_k + x_m - 2y\|^2 \\ &= 2\|x_k - y\|^2 + 2\|x_m - y\|^2 - 4\|\frac{x_k + x_m}{2} - y\|^2 \leq 2\|x_k - y\|^2 + 2\|x_m - y\|^2 \\ &\quad - 4\delta^2 \rightarrow 2\delta^2 + 2\delta^2 - 4\delta^2 = 0 \text{ cuando } k, m \rightarrow \infty. \end{aligned}$$

Note que $\frac{x_k + x_m}{2} \in C$ y $\{x_k\}_{k \in \mathbb{N}}$ es una sucesión de Cauchy. Por lo tanto $\{x_k\}_{k \in \mathbb{N}}$ tiene un límite $x^* \in C$, ya que, C es un conjunto cerrado, así, $\lim_{k \rightarrow \infty} x_k = x^* \in C$. Para probar la unicidad, suponga que $\lim_{k \rightarrow \infty} x_k = \bar{x} \in C$, entonces $0 \leq \|x^* - \bar{x}\|^2 \leq 2\|x_k - x^*\|^2 + 2\|x_k - \bar{x}\|^2 - 4\delta^2 = 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$, de donde $x^* = \bar{x}$.

Suponga ahora que $\langle x - x^*, x^* - y \rangle \geq 0$ para toda $x \in C$, entonces $\|y - x\|^2 = \|y - x^* + (x^* - x)\|^2 = \|x - x^*\|^2 + \|x^* - y\|^2 + 2\langle x - x^*, x^* - y \rangle \geq \|x^* - y\|^2$ para toda $x \in C$.

Recíprocamente, si $\|y - x^*\|^2 \leq \|y - x\|^2$ para toda $x \in C$, entonces

si $\lambda > 0$, $\|y - (\lambda x + (1-\lambda)x^*)\|^2 = \|y - x^* - \lambda(x - x^*)\|^2 \geq \|y - x^*\|^2$ para toda $x \in C$. Por otra parte $\|y - x^* - \lambda(x - x^*)\|^2 = \|y - x^*\|^2 + \lambda^2 \|x - x^*\|^2 + 2\lambda \langle x - x^*, x^* - y \rangle$, lo cual implica $\lambda^2 \|x - x^*\|^2 + 2\lambda \langle x - x^*, x^* - y \rangle \geq 0$ para toda $x \in C$, así dividiendo por $\lambda > 0$ se tiene $\lambda \|x - x^*\|^2 + 2 \langle x - x^*, x^* - y \rangle \geq 0$ para toda $x \in C$. Si $\lambda \rightarrow 0$, se sigue que $\langle x - x^*, x^* - y \rangle \geq 0$ para todo $x \in C$.

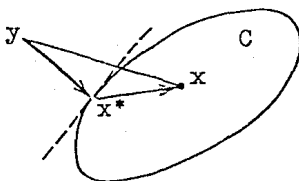


Figura 11. Distancia mínima de $y \notin C$ a C .

Uno de los conjuntos convexos más importante es el hiperplano. El concepto de hiperplano es fundamental en el entendimiento de la teoría de optimización. La definición más natural de hiperplano se obtiene de la generalización lógica de las propiedades geométricas de un plano en \mathbb{R}^3 .

Definición. 1.1.2.26 Un conjunto $V \subset \mathbb{R}^n$ es llamado una variedad lineal si dados $x, y \in V$, se tiene que $\lambda x + (1-\lambda)y \in V$ para todo $\lambda \in \mathbb{R}$.

Observe que la única diferencia de esta definición con la de conjunto convexo, es que, dados dos puntos del conjunto, toda la línea que pasa a través de ellos pertenece también al conjunto. En \mathbb{R}^n todos los subespacios trasladados son variedades

lineales, en general, se puede hablar de la dimensión de una variedad como la dimensión del subespacio, que trasladado coincide con la variedad.

Definición. 1.1.2.27 Un hiperplano en \mathbb{R}^n , es una variedad de dimensión $n-1$.

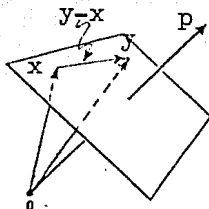


Figura 12. Un hiperplano en el espacio \mathbb{R}^3 .

Lema. 1.1.2.28 Sea $V \subset \mathbb{R}^n$ una variedad lineal y sea $y \in V$. Si $M = V - y$, entonces M es un subespacio de \mathbb{R}^n .

Prueba:

Claramente $0 \in M$. Sea $x \in M$ y $\alpha \in \mathbb{R}$, entonces $x = v - y$ y $\alpha x = \alpha v - \alpha y + y - y = \alpha v + (1 - \alpha)y - y$ con $\alpha v + (1 - \alpha)y \in V$, de donde $\alpha x \in M$. Sean $x, z \in M$, entonces $x = v - y, z = w - y$ con $v, w \in V$, entonces $x + y = v + w - 2y$, de donde $\frac{1}{2}(x + y) = \frac{1}{2}(v + w) - y$, siendo así que $\frac{1}{2}(x + y) \in M$, de donde $x + y \in M$.

Proposición. 1.1.2.29 Un subconjunto H de \mathbb{R}^n , es un hiperplano si, y sólo si existen $p \in \mathbb{R}^n, p \neq 0$ y $\alpha \in \mathbb{R}$, tales que $H = \{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$.

Prueba:

Suponga que existen $p \in \mathbb{R}^n, p \neq 0$ y $\alpha \in \mathbb{R}$, tales que $H = \{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$. Observe que H es una variedad lineal, ya que, si $x, y \in H$ y $\lambda \in \mathbb{R}$, entonces $\langle p, \lambda x + (1 - \lambda)y \rangle = \lambda \langle p, x \rangle + (1 - \lambda) \langle p, y \rangle = \lambda \alpha + (1 - \lambda) \alpha = \alpha$. Sea $y \in H$, entonces por el lema 1.1.2.28 $M = H - y$ es un subespacio.

Este subespacio consiste de todos aquellos elementos $z \in \mathbb{R}^n$ tales que $\langle p, z \rangle = 0$, ya que, si $z \in M$, entonces $z = h - y$ con $h \in H$, de donde $\langle p, z \rangle = \langle p, h \rangle - \langle p, y \rangle = \alpha - \alpha = 0$ y por lo tanto $\dim M = n - 1$ puesto que $M^\perp = \langle p \rangle$, donde $\langle p \rangle$ denota el subespacio generado por el vector p .

Recíprocamente, si el conjunto H es un hiperplano, entonces $M = H - y$ es un subespacio de dimensión $n - 1$. Sea $p \in \mathbb{R}^n$, $p \neq 0$, tal que $M^\perp = \langle p \rangle$, así $M = \{x \in \mathbb{R}^n; \langle p, x \rangle = 0\}$. Tome $\alpha = \langle p, y \rangle$, si $x \in H$, entonces $x - y \in H - y$, teniéndose que $\langle p, x \rangle - \langle p, y \rangle = 0$, de donde $\langle p, x \rangle = \alpha$. Por lo tanto $H \subset \{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$ y como ya se sabe que $\{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$ tiene dimensión $n - 1$ al igual que H , se concluye con esto la prueba de la proposición establecida.

Por la proposición anterior se sigue que un hiperplano H en \mathbb{R}^n , es un conjunto de la forma $\{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$, donde $p \in \mathbb{R}^n$, $p \neq 0$ y $\alpha \in \mathbb{R}$. El vector p es llamado el vector normal a H . Un hiperplano H define dos semiespacios cerrados (en el sentido topológico) $H(\geq) = \{x \in \mathbb{R}^n; \langle p, x \rangle \geq \alpha\}$ y $H(\leq) = \{x \in \mathbb{R}^n; \langle p, x \rangle \leq \alpha\}$. Asimismo, H define dos semiespacios abiertos (en el sentido topológico) $H(>) = \{x \in \mathbb{R}^n; \langle p, x \rangle > \alpha\}$ y $H(<) = \{x \in \mathbb{R}^n; \langle p, x \rangle < \alpha\}$. Se concluye entonces que cualquier punto $x \in \mathbb{R}^n$ satisface la tricotomía $x \in H(>)$ ó $x \in H(<)$ ó $x \in H$.

También un hiperplano H puede ser referido a uno de sus puntos, esto es, si $y \in H$, entonces $\langle p, y \rangle = \alpha$ y en consecuencia $H = \{x \in \mathbb{R}^n; \langle p, x - y \rangle = 0\}$.

Definición. 1.1.2.30 Sean $S, T \subset \mathbb{R}^n$ dos conjuntos no vacíos y sea $H = \{x \in \mathbb{R}^n; \langle p, x \rangle = \alpha\}$ un hiperplano.

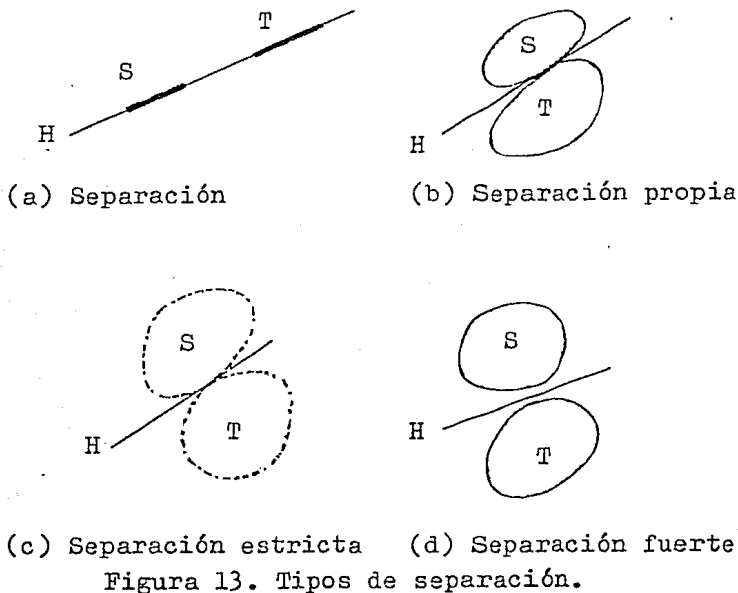
(i). Se dice que el hiperplano H separa a S de T , si $\langle p, x \rangle \geq \alpha$ para toda $x \in S$ y $\langle p, x \rangle \leq \alpha$ para toda $x \in T$.

(ii). Si en (i) se tiene que $S \cup T \neq H$, entonces se dice que H separa propiamente a S de T .

(iii). Se dice que el hiperplano H separa estrictamente a S de T , si $\langle p, x \rangle > \alpha$ para toda $x \in S$ y $\langle p, x \rangle < \alpha$ para toda $x \in T$.

(iv). Se dice que el hiperplano H separa fuertemente a S de T si existe un número $\epsilon > 0$, tal que $\langle p, x \rangle \geq \alpha + \epsilon$ para toda $x \in S$ y $\langle p, x \rangle \leq \alpha$ para toda $x \in T$.

Una interpretación geométrica de la definición anterior se muestra en la siguiente figura:



Obviamente se tiene:

Proposición. 1.1.2.31 Separación fuerte \Rightarrow Separación estricta
 \Rightarrow Separación propia \Rightarrow Separación.

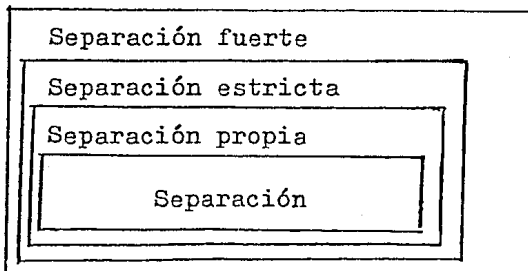


Figura 14. Contención entre separaciones.

A continuación se presentan los teoremas de separación:

Teorema 1.1.2.32 Sea S un conjunto convexo, no vacío y cerrado de \mathbb{R}^n y sea $y \notin S$. Entonces existen un vector no nulo $p \in \mathbb{R}^n$ y un escalar $\alpha \in \mathbb{R}$, tales que $\langle p, y \rangle > \alpha$ y $\langle p, x \rangle \leq \alpha$ para todo $x \in S$ y por lo tanto la separación es fuerte.

Prueba:

Puesto que S es un conjunto convexo, no vacío y cerrado de \mathbb{R}^n y $y \notin S$. Por el teorema de proyección, existe un único $x^* \in S$ tal que $\langle x - x^*, y - x^* \rangle \leq 0$ para todo $x \in S$, equivalentemente $\langle x, y - x^* \rangle \leq \langle x^*, y - x^* \rangle$ para todo $x \in S$, entonces $\|y - x^*\|^2 = \langle y - x^*, y - x^* \rangle = \langle y, y - x^* \rangle - \langle x^*, y - x^* \rangle \leq \langle y, y - x^* \rangle - \langle x, y - x^* \rangle = \langle y - x, y - x^* \rangle$ para toda $x \in S$. Sea $p = y - x^* \neq 0$, así $\|y - x^*\|^2 \leq \langle y - x, p \rangle$ y por lo tanto $\langle p, x \rangle + \|y - x^*\|^2 \leq \langle p, y \rangle$ para todo $x \in S$ ó $\langle p, x \rangle \leq \langle p, y \rangle$ para todo $x \in S$. Sea $\alpha = \sup\{\langle p, x \rangle; x \in S\}$, entonces $\langle p, x \rangle \leq \alpha$ para todo $x \in S$ y $\langle p, y \rangle > \alpha$.

Corolario. 1.1.2.33 Sea S un conjunto convexo, no vacío y cerrado de \mathbb{R}^n . Entonces S es la intersección de todos los semiespacios que lo contienen.

Prueba:

Obviamente S está contenido en la intersección de todos los semiespacios que lo contienen. Para verificar la otra contención. Suponga que por el contrario, que existe un $y \notin S$ que pertenece a la intersección de todos los semiespacios que contienen a S . Por el teorema 1.1.2.32, existe un semiespacio que contiene a S pero no a y . De donde y no está en la intersección de todos los semiespacios que contienen a S , lo cual es una contradicción, probándose así el corolario propuesto.

La discusión próxima será basada en el teorema de Farkas, el cual ha sido ampliamente usado en la obtención de condiciones de optimalidad de problemas de programación ya sea lineal o no lineal.

Teorema 1.1.2.34 (Lema de alternativa de Farkas)

Sea A una matriz de orden $m \times n$ y c un vector de \mathbb{R}^n . Entonces uno y sólo uno de los siguientes dos sistemas tiene solución:

Sistema (i). $Ax \leq 0$ y $\langle c, x \rangle > 0$ para algún $x \in \mathbb{R}^n$

Sistema (ii). $A^t y = c$ y $y \geq 0$ para algún $y \in \mathbb{R}^m$

Prueba:

Suponga que el sistema (ii) tiene solución; esto es, existe $y \in \mathbb{R}^m$, $y \geq 0$, tal que $A^t y = c$, entonces $0 < \langle c, x \rangle = y^t A x \leq 0$. Por lo tanto el sistema (i) no tiene solución.

Ahora suponga que (ii) no tiene solución. Sea $S = \{x \in \mathbb{R}^n; x = A^t y, y \geq 0\}$. Note que el conjunto S es convexo y cerrado y que $c \notin S$. Por el teorema 1.1.2.32 existen $p \in \mathbb{R}^n$, $p \neq 0$ y un escalar $\alpha \in \mathbb{R}$, tales que $\langle p, c \rangle > \alpha$ y $\langle p, x \rangle \leq \alpha$ para todo $x \in S$. Es claro que $0 \in S$, así $0 = \langle p, 0 \rangle \leq \alpha$, teniéndose pues que $\langle p, c \rangle > 0$. También $\alpha \geq \langle p, x \rangle = p^t A^t y = (Ap)^t y = y^t (Ap)$ para todo $y \geq 0$, pero como y puede ser tan grande como se desee, se implica que $Ap \leq 0$, es decir existe un vector $p \in \mathbb{R}^n$, tal que $\langle p, x \rangle > 0$ y $Ap \leq 0$, siendo así $p \in \mathbb{R}^n$ una solución del sistema (i).

El teorema anterior acepta una interpretación geométrica sencilla: Sean q_1, q_2, \dots, q_m las columnas de la matriz A^t , entonces el sistema (ii) tiene solución si c pertenece al cono convexo generado por q_1, q_2, \dots, q_m y el sistema (i) tiene solución si el cono convexo cerrado $\{x \in \mathbb{R}^n; Ax \leq 0\}$ y el semiespacio abierto $\{x \in \mathbb{R}^n; \langle c, x \rangle > 0\}$ tienen intersección no vacía.

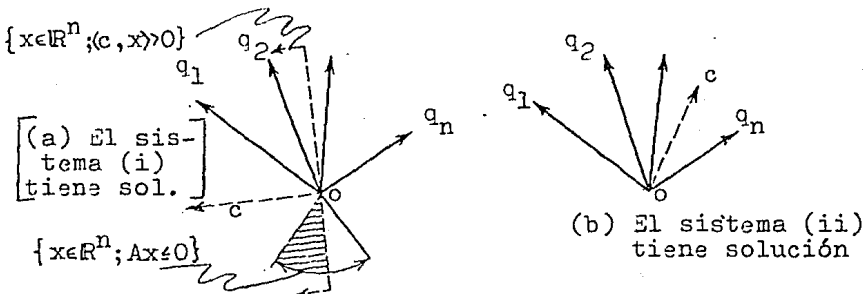


Figura 15. Interpretación geométrica del teorema de Farkas.

Hasta el momento se ha estudiado sobre la separación entre un conjunto convexo cerrado y puntos $y \in \mathbb{R}^n$ que no están en él. Si S es abierto o no es cerrado, entonces pueden suceder dos cosas: A saber,

- (i) $y \notin \bar{S}$, en cuyo caso se aplica el teorema 1.1.2.32
- (ii) $y \in \partial S$

Se verá a continuación que para el caso (ii), un conjunto convexo S tiene un hiperplano soporte en cada punto de ∂S en el siguiente sentido:

Definición. 1.1.2.35 Sea S un subconjunto no vacío de \mathbb{R}^n (S , no necesariamente un conjunto convexo) y sea $\bar{x} \in \partial S$. Se dice que un hiperplano $H = \{x \in \mathbb{R}^n; \langle p, x - \bar{x} \rangle = 0\}$ soporta a S en \bar{x} si, una y sólo una de las siguientes condiciones se cumple:

- (i). $SCH(\geq)$, esto es, $\langle p, x - \bar{x} \rangle \geq 0$ para toda $x \in S$
- (ii). $SCH(\leq)$, es decir, $\langle p, x - \bar{x} \rangle \leq 0$ para toda $x \in S$.

Si en adición, en cualquier caso, $S \cap H = \{\bar{x}\}$, se dice entonces que el hiperplano H soporta propiamente a S en \bar{x} .

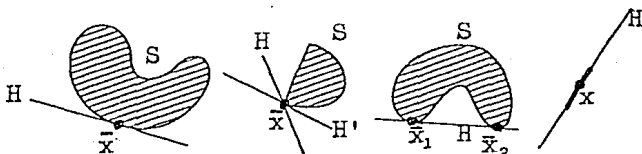
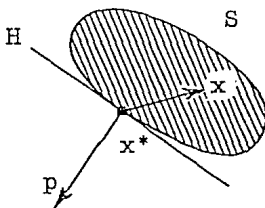


Figura 16. Ejemplos de hiperplanos de soporte

Teorema. 1.1.2.36 Sea S un subconjunto convexo y no vacío de \mathbb{R}^n (S , no necesariamente cerrado) y sea $\bar{x} \in \partial S$. Entonces existe un hiperplano que soporta a S en \bar{x} . Equivalentemente, existe un vector no nulo $p \in \mathbb{R}^n$, tal que $\langle p, x - \bar{x} \rangle \leq 0$ para toda $x \in S$.

Prueba:

Sea $\bar{x} \in \partial S$ y sea $\{y_k\}_{k \in \mathbb{N}}$ una sucesión tal que $y_k \notin S$ para toda $k \in \mathbb{N}$ y $y_k \rightarrow \bar{x}$ cuando $k \rightarrow \infty$. Por el teorema 1.1.2.32, para cada y_k existe un p_k , el cual sin pérdida de generalidad satisface $\|p_k\|=1$, tal que $\langle p_k, y_k \rangle > \langle p_k, x \rangle$ para toda $x \in S$. Puesto que la sucesión $\{y_k\}_{k \in \mathbb{N}}$ está contenida en el conjunto compacto $\{x \in \mathbb{R}^n; \|x\|=1\}$, se tiene que existe una subsucesión $\{y_{k_m}\}_{m \in \mathbb{N}}$, la cual converge a un elemento $p \in \{x \in \mathbb{R}^n; \|x\|=1\}$, note que $1 = \lim_{m \rightarrow \infty} \|p_{k_m}\| = \|\lim_{m \rightarrow \infty} p_{k_m}\| = \|p\|$. Ahora bien, como $\langle p_{k_m}, y_{k_m} \rangle > \langle p_{k_m}, x \rangle$ para toda $x \in S$, entonces $\lim_{m \rightarrow \infty} \langle p_{k_m}, y_{k_m} \rangle \geq \lim_{m \rightarrow \infty} \langle p_{k_m}, x \rangle$ para toda $x \in S$, y por la continuidad del producto escalar se sigue que $\langle p, x \rangle - \langle p, \bar{x} \rangle = \langle p, x - \bar{x} \rangle \leq 0$ para toda $x \in S$.



Figural7. Interpretación geométrica del teorema 1.1.2.36

Hasta aquí, se han presentado resultados sobre la separación por un hiperplano entre un conjunto convexo y un punto que no pertenece al conjunto y sobre hiperplanos que soportan a conjuntos convexos en puntos de la frontera. A continuación

se establecen algunos resultados sobre la separación de conjuntos convexos, ajenos y no vacíos por un hiperplano, de tal manera que uno de ellos pertenece a $H(\geq)$ y el otro a $H(\leq)$. Más precisamente:

Teorema. 1.1.2.37 Sean S y T dos subconjuntos convexos, ajenos y no vacíos de \mathbb{R}^n . Entonces existe un hiperplano H que separa a S de T ; es decir, existe $p \in \mathbb{R}^n$, $p \neq 0$, tal que $\sup\{\langle p, x \rangle; x \in S\} \leq \inf\{\langle p, z \rangle; z \in T\}$.

Prueba:

Defínase el conjunto $Q = S - T = \{y \in \mathbb{R}^n; y = x - z, x \in S \text{ y } z \in T\}$, Q es claramente convexo. Ahora bien $0 \notin Q$, ya que $S \cap T = \emptyset$. Por el teorema 1.1.2.32, existen $p \in \mathbb{R}^n$, $p \neq 0$ y un escalar $\alpha \in \mathbb{R}$, tales que $\langle p, y \rangle \leq \alpha$ para toda $y \in Q$ y $\langle p, 0 \rangle > \alpha$, teniendo así que $\langle p, y \rangle < 0$ para toda $y \in Q$, es decir, $\langle p, x \rangle < \langle p, z \rangle$ para todo $x \in S$ y para todo $z \in T$, y así $\sup\{\langle p, x \rangle; x \in S\} \leq \inf\{\langle p, z \rangle; z \in T\}$.

Esta sección se concluye con otro resultado importante en la teoría de desigualdades.

Teorema 1.1.2.38 (Teorema de alternativa de Gordan)

Sea A una matriz de $m \times n$. Entonces uno y sólo uno de los dos siguientes sistemas tiene solución:

Sistema (i). $Ax < 0$ para algún $x \in \mathbb{R}^n$

Sistema (ii). $A^t p = 0$ y $p \geq 0$, $p \neq 0$ para algún $p \in \mathbb{R}^m$.

Prueba:

Primero se verá que si el sistema (i) tiene solución, entonces el sistema (ii) no tiene solución. Suponga que por el contrario, que los sistemas (i) y (ii) tienen soluciones \hat{x} y \hat{p} , entonces $A\hat{x} < 0$ y como $\hat{p} \geq 0$, $\hat{p} \neq 0$, se tiene $0 > \hat{p}^t A\hat{x} = (A^t \hat{p})^t \hat{x} = 0$, lo cual es una contradicción. Por lo tanto, si el sistema (i), tiene solución, entonces el sistema (ii) no tiene solución. Suponga ahora que el sistema (i) no tiene solución y considere

los dos siguientes conjuntos convexos, ajenos y no vacíos $S = \{z \in \mathbb{R}^m; z = Ax, x \in \mathbb{R}^n\}$ y $T = \{z \in \mathbb{R}^m; z < 0\}$. Por el teorema anterior, existe un vector $p \in \mathbb{R}^m$, $p \neq 0$, tal que $\langle p, Ax \rangle \geq \langle p, z \rangle$ para todo $x \in \mathbb{R}^n$ y para todo $z \in \bar{T}$. Puesto que el vector $z \in T$, $z < 0$, puede hacerse tan grande en valor absoluto como se quiera, se sigue que $p \geq 0$, y como $0 \in \bar{T}$, se sigue también que $p^t Ax \geq 0$ para toda $x \in \mathbb{R}^n$, escogiendo $x = -A^t p$, se tiene $-\|A^t p\|^2 \geq 0$, de donde $A^t p = 0$. Por lo tanto el sistema (ii) tiene una solución.

1.1.3 FUNCIONES CONVEXAS Y SU GENERALIZACION

En esta sección se introduce el concepto de función convexa y se discute sobre su propiedad de continuidad en el interior de su dominio. A continuación se presenta el concepto de subgradiente, el cual recientemente ha despertado un creciente interés en la programación no lineal. Este concepto está relacionado con el siguiente hecho geométrico: El epígrafe de una función convexa, es un conjunto convexo y por lo tanto tiene un hiperplano de soporte en cada punto de su frontera. Después se presenta la caracterización de las funciones convexas diferenciables, así como un conjunto de condiciones necesarias y suficientes de soluciones óptimas de problemas de programación no lineal convexa. Por último se presentan las generalizaciones más importantes de las funciones convexas.

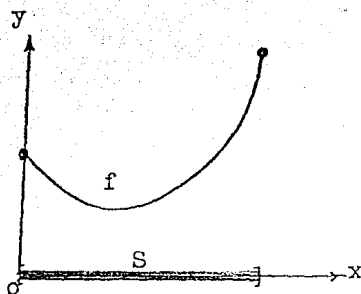
En lo que sigue, se establecen las propiedades más importantes que poseen las funciones convexas:

Definición. 1.1.3.1 Sea $f: S \rightarrow \mathbb{R}$, $S \subset \mathbb{R}^n$, una función definida sobre un conjunto convexo y no vacío S . Se dice que f es una función convexa, si dados $x, y \in S$, se tiene que $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ para todo $\lambda \in [0, 1]$.

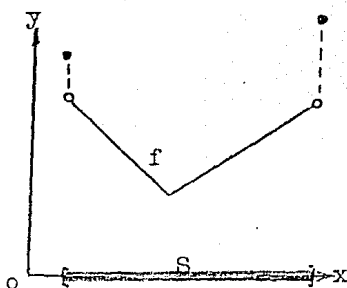
Se dice que una función $f: S \rightarrow \mathbb{R}$ es cóncava si $-f$ es convexa.

Un par de resultados inmediatos que se desprenden de la definición anterior son que: Si $f: S \rightarrow \mathbb{R}$, $S \subset \mathbb{R}^n$ es una función convexa, entonces el conjunto de nivel α , $S_\alpha = \{x \in S; f(x) \leq \alpha\}$ es un conjunto convexo para toda $\alpha \in \mathbb{R}$, y que si $\{x_1, x_2, \dots, x_k\} \subset S$ y $\{\lambda_1, \lambda_2, \dots, \lambda_k\} \subset \mathbb{R}$ son tales que $\lambda_m \geq 0$ para toda m , $\sum_{m=1}^k \lambda_m = 1$, entonces $f(\sum_{m=1}^k \lambda_m x_m) \leq \sum_{m=1}^k \lambda_m f(x_m)$.

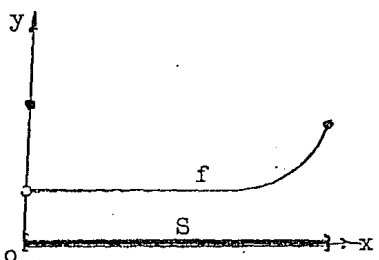
La noción de función convexa es gráficamente representada en la siguiente figura:



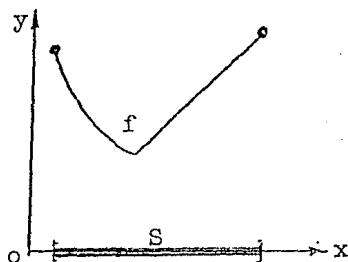
(a) función convexa



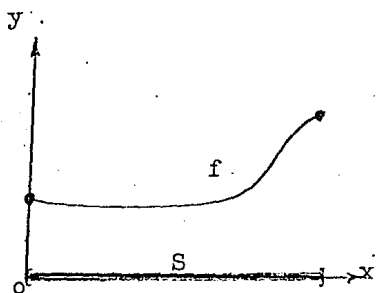
(b) función convexa



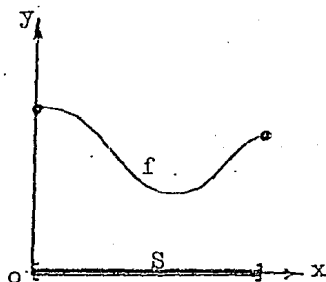
(c) función convexa



(d) función convexa



(e) función no convexa



(f) función no convexa

Figura 18. Ejemplos de funciones convexas y no convexas

De las propiedades relevantes que se estudiarán, la primera consiste en el hecho de que, si $f: S \rightarrow \mathbb{R}$ es una función convexa definida sobre un conjunto convexo SCR^n , entonces f es una función continua en el interior de su dominio.

Definición. 1.1.3.2 Sea $f: S \rightarrow \mathbb{R}$, SCR^n , $S \neq \emptyset$, se dice que la función f es continua en el punto $x \in S$ si, dado $\epsilon > 0$, existe $\delta > 0$ tal que $\|x-y\| \leq \delta$ implica $|f(x)-f(y)| \leq \epsilon$.

Si f es continua en todo punto de S , se dice entonces que f es continua en S .

La definición anterior expresa que para cada vecindad V de $f(x)$, existe una vecindad W de x , tal que $f(y) \in V$ para toda $y \in W$, es decir, $f(W) \subset V$.

Un hecho útil y sencillo de probar, es que: Una función $f: S \rightarrow \mathbb{R}$, SCR^n , es continua en el punto $x \in S$ si, y sólo si $f(x_k) \rightarrow f(x)$ para toda sucesión $\{x_k\}_{k \in \mathbb{N}}$, tal que $x_k \rightarrow x$ cuando $k \rightarrow \infty$.

Teorema. 1.1.3.3 Sea S un subconjunto convexo no vacío de \mathbb{R}^n con $\text{Int}(S) \neq \emptyset$ y sea $f: S \rightarrow \mathbb{R}$ una función convexa. Entonces f es continua en $\text{Int}(S)$.

Prueba:

Sea $y \in \text{Int}(S)$, se verá que f es continua en y , es decir, que dado $\epsilon > 0$, existe $\delta > 0$, tal que, $\|w-y\| \leq \delta$ implica $|f(w)-f(y)| \leq \epsilon$.
 Sea pues, $\epsilon > 0$, puesto que $y \in \text{Int}(S)$, se sigue que existe $\delta' > 0$, tal que $\|x-y\| < \delta'$ implica $x \in S$. Tome cualquier $w \in \{x \in \mathbb{R}^n; \|x-y\| < \frac{\delta'}{\sqrt{n}}\}$.
 Si $w_k - y_k \geq 0$, sea $z_k = \delta' e_k$, en caso contrario $z_k = -\delta' e_k$, donde $\{e_k\}_{1 \leq k \leq n}$ es la base algebraica canónica de \mathbb{R}^n . Entonces $w-y = \sum_{k=1}^n \alpha_k z_k$ donde $\alpha_k \geq 0$ para toda k . En efecto, si $\{e_k\}_{1 \leq k \leq n}$ es base algebraica de \mathbb{R}^n , entonces también lo es $\{\delta' e_k\}_{1 \leq k \leq n}$. Algo más, si cualquier vector $\delta' e_k$ se reemplaza por $-\delta' e_k$, entonces también se tendrá que $\{z_k\}_{1 \leq k \leq n}$ es una base algebraica de \mathbb{R}^n . Por lo tanto, existen escalares $\alpha_k \in \mathbb{R}$ ($k=1, 2, \dots, n$),

tales que $\alpha_k \geq 0$ para toda k y $w-y = \sum_{k=1}^n \alpha_k z_k$. En consecuencia $\|w-y\|^2 = \sum_{k=1}^n \delta^2 \alpha_k^2$, de donde $\|w-y\| = \delta \left[\sum_{k=1}^n \alpha_k^2 \right]^{\frac{1}{2}}$. Ahora examínese la

función f evaluada en el punto w ,

$$\begin{aligned} f(w) &= f\left(y + \sum_{k=1}^n \alpha_k z_k\right) = f\left(\sum_{k=1}^n \frac{1}{n} (y + n\alpha_k z_k)\right) \\ &\leq \sum_{k=1}^n \frac{1}{n} f(y + n\alpha_k z_k) \\ &= \frac{1}{n} \sum_{k=1}^n f(y - n\alpha_k y + n\alpha_k y + n\alpha_k z_k) \\ &= \frac{1}{n} \sum_{k=1}^n f((1 - n\alpha_k)y + n\alpha_k(y + z_k)) \end{aligned}$$

Si se quiere aplicar la propiedad de convexidad de la función f como sigue:

$$f((1 - n\alpha_k)y + n\alpha_k(y + z_k)) \leq (1 - n\alpha_k)f(y) + n\alpha_k f(y + z_k),$$

se requiere que $0 \leq n\alpha_k \leq 1$, en cuyo caso se debe pedir que $\alpha_k \leq \frac{1}{n}$, así $\|w-y\| = \delta \left(\sum_{k=1}^n \alpha_k^2 \right)^{\frac{1}{2}} \leq \frac{\delta'}{\sqrt{n}}$, y por lo tanto

$$\begin{aligned} f(w) &\leq \frac{1}{n} \sum_{k=1}^n ((1 - n\alpha_k)f(y) + n\alpha_k f(y + z_k)) = f(y) + \sum_{k=1}^n \alpha_k (f(y + z_k) - f(y)), \\ \text{de donde } f(w) - f(y) &\leq \sum_{k=1}^n \alpha_k (f(y + z_k) - f(y)). \end{aligned}$$

Si se quiere ahora que $f(w) - f(y) \leq \epsilon$, ponga $\theta = \max_{1 \leq k \leq n} \{\max\{f(y + \delta e_k) - f(y), f(y - \delta e_k) - f(y)\}\}$. Observe que θ es un número finito no negativo, si $\theta = 0$ reemplace θ por $\theta + 1$. Entonces $f(y + z_k) - f(y) \leq \theta$ para toda k , teniéndose que $f(w) - f(y) \leq \theta \sum_{k=1}^n \alpha_k$. Si se toman los α_k , de tal manera que $\alpha_k \leq \frac{\epsilon}{n\theta}$ para todo k , entonces $\|w-y\| \leq \delta$ implica que $f(w) - f(y) \leq \epsilon$, donde $\delta = \min\left\{\frac{\delta'}{\sqrt{n}}, \frac{\delta' \epsilon}{\sqrt{n}\theta}\right\}$.

Para concluir la prueba del teorema, basta demostrar que $f(y) - f(w) \leq \epsilon$. Sea $z = 2y - w$, entonces $\|z-y\| = \|y-w\| \leq \delta$, en consecuencia $f(z) - f(y) \leq \epsilon$, pero $y = \frac{1}{2}z + \frac{1}{2}w$, y puesto que f es una función convexa $f(y) \leq \frac{1}{2}f(z) + \frac{1}{2}f(w)$ ó $f(y) - f(w) \leq f(z) - f(y) \leq \epsilon$.

Uno de los conceptos particularmente útil en el desarrollo de condiciones necesarias de optimalidad en programación no lineal, es el concepto de razón de cambio instantáneo del valor de una función en una dirección.

Definición. 1.1.3.4 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$ y abierto y sea $f: S \rightarrow \mathbb{R}$. Sea $x \in S$ y $d \in \mathbb{R}^n$, $d \neq 0$, tal que $x + \lambda d \in S$ para $\lambda > 0$ suficientemente pequeño. La derivada de Gateaux de la función f en el punto $x \in S$ en la dirección $d \in \mathbb{R}^n$, denotada por $Df(x; d)$ está dada por el siguiente límite, si es que existe:

$$Df(x; d) = \lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda} = \left. \frac{d}{d\lambda} f(x + \lambda d) \right|_{\lambda=0}$$

Si la función f es Gateaux-diferenciable en todo punto de S , se dice entonces que f es Gateaux-diferenciable en S .

Una propiedad importante de las funciones convexas Gateaux-diferenciables es mostrada en la siguiente proposición:

Proposición. 1.1.3.5 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$ abierto y convexo y sea $f: S \rightarrow \mathbb{R}$ una función convexa Gateaux-diferenciable en el punto $x \in S$, entonces $Df(x; d) = \inf \left\{ \frac{f(x + \lambda d) - f(x)}{\lambda} ; \lambda > 0 \right\}$, donde $d \in \mathbb{R}^n$, $d \neq 0$ y $\lambda > 0$ suficientemente pequeña, tal que $x + \lambda d \in S$.

Prueba:

Sean $\mu, \lambda \in \mathbb{R}$, $\mu > \lambda > 0$, suficientemente pequeños. Por ser f una función convexa, se tiene que $f(x + \lambda d) = f\left(\frac{\lambda}{\mu}(x + \mu d) + \left(1 - \frac{\lambda}{\mu}\right)x\right) \leq \frac{\lambda}{\mu} f(x + \mu d) + \left(1 - \frac{\lambda}{\mu}\right) f(x)$, ésta última desigualdad implica $\frac{f(x + \lambda d) - f(x)}{\lambda} \leq \frac{f(x + \mu d) - f(x)}{\mu}$. Sea $\varphi: (0, \infty) \rightarrow \mathbb{R}$, definida por $\varphi(\lambda) = \frac{f(x + \lambda d) - f(x)}{\lambda}$, por lo tanto si $\lambda < \mu$, entonces $\varphi(\lambda) \leq \varphi(\mu)$, es decir φ es una función no decreciente de $\lambda > 0$ y acotada inferiormente por $Df(x; d)$ con lo que se concluye la prueba de la proposición.

A continuación se estudia el concepto de subgradiente de una función convexa, ésto será hecho estableciendo primero los conceptos de epígrafe e hipógrafe de una función convexa, los cuales a su vez, como veremos, son útiles en la caracterización de las funciones convexas.

Una función $f: S \rightarrow \mathbb{R}^n$ definida sobre un subconjunto no vacío S de \mathbb{R}^n puede ser totalmente descrita mediante el conjunto $\{(x, f(x)); x \in S\} \subset \mathbb{R}^{n+1}$, este conjunto es llamado la gráfica de f y es denotado por $\text{graf}(f)$. Existen dos conjuntos estrechamente ligados a $\text{graf}(f)$: la epígrafe y la hipógrafe denotadas, respectivamente por $\text{epí}(f)$ e $\text{hip}(f)$.

Definición. 1.1.3.6 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$ y sea $f: S \rightarrow \mathbb{R}$, entonces

(i). La epígrafe de la función f , consiste en el conjunto $\text{epí}(f) = \{(\gamma, f(x)); f(x) \leq \gamma, x \in S, \gamma \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$.

(ii). La hipógrafe de la función f , consiste en el conjunto $\text{hip}(f) = \{(\gamma, f(x)); f(x) \geq \gamma, x \in S, \gamma \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$.

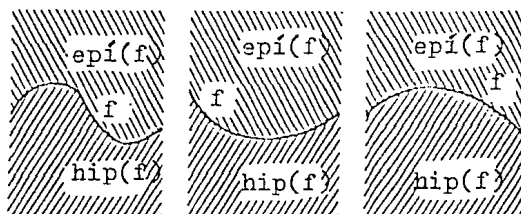


Figura 19. Ejemplos de epígrafes e hipógrafes.

El conjunto $\text{epí}(f)$ consiste de aquellos puntos en \mathbb{R}^{n+1} que están por encima de $\text{graf}(f)$ y el conjunto $\text{hip}(f)$ consiste de aquellos puntos en \mathbb{R}^{n+1} que están por debajo de $\text{graf}(f)$. Estas nociones conducen a una caracterización muy importante de las funciones convexas.

Proposición. 1.1.3.7 Sea S un subconjunto convexo y no vacío de \mathbb{R}^n y $f: S \rightarrow \mathbb{R}$ una función. Entonces, f es una función convexa si, y sólo si su epígrafe es un conjunto convexo.

Prueba:

Suponga primero que f es una función convexa y sean $(x, y_1), (y, y_2) \in \text{epí}(f)$. Se tiene que ver que $\lambda(x, y_1) + (1-\lambda)(y, y_2) \in \text{epí}(f)$ para toda $\lambda \in [0, 1]$. Si $(x, y_1), (y, y_2) \in \text{epí}(f)$, entonces $f(x) \leq y_1$ y $f(y) \leq y_2$. Sea $\lambda \in [0, 1]$, entonces $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda y_1 + (1-\lambda)y_2$ y por ser S un conjunto convexo $\lambda x + (1-\lambda)y \in S$ y lo anterior implica que $(\lambda x + (1-\lambda)y, \lambda y_1 + (1-\lambda)y_2) \in \text{epí}(f)$ para toda $\lambda \in [0, 1]$, es decir, $\text{epí}(f)$ es un conjunto convexo.

Recíprocamente, suponga que $\text{epí}(f)$ es un conjunto convexo y sean $x, y \in S$, entonces $(x, f(x)), (y, f(y)) \in \text{epí}(f)$ y por la convexidad de $\text{epí}(f)$, se sigue que $(\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda)f(y)) \in \text{epí}(f)$ para todo $\lambda \in [0, 1]$, por definición lo anterior implica $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ para toda $\lambda \in [0, 1]$, lo cual implica que f es una función convexa.

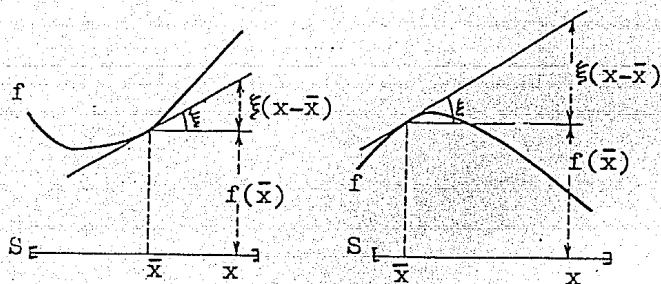
De manera análoga, se puede demostrar que el hipógrafe de una función cóncava es un conjunto convexo e inversamente.

En el transcurso de la sección anterior se estudió sobre la existencia de hiperplanos que soportan conjuntos convexos en puntos de su frontera. Este hecho conduce a la noción de subgradiente, el cual se discute a continuación:

Definición. 1.1.3.8

(i). Sea $S \neq \emptyset$ un subconjunto convexo de \mathbb{R}^n y sea $f: S \rightarrow \mathbb{R}$ una función convexa. Se dice que el vector ξ es un subgradiente de f en el punto $\bar{x} \in S$ si $f(x) \geq f(\bar{x}) + \langle \xi, x - \bar{x} \rangle$ para toda $x \in S$.

(ii). Análogamente, si f es una función cóncava, $\xi \in \mathbb{R}^n$ es un subgradiente de f en $\bar{x} \in S$ si $f(x) \leq f(\bar{x}) + \langle \xi, x - \bar{x} \rangle$ para toda $x \in S$.



(a) Subgradiente de una función convexa

(b) Subgradiente de una función cóncava

Figura 20. Ejemplos de subgradietes

Teorema. 1.1.3.9 Sea S un subconjunto convexo con interior no vacío en \mathbb{R}^n y sea $f: S \rightarrow \mathbb{R}$ una función convexa sobre S . Entonces para cada $\bar{x} \in \text{Int}(S)$, existe un subgradiente ξ de f , tal que el hiperplano $H = \{(x, y); y = f(\bar{x}) + \langle \xi, x - \bar{x} \rangle, x \in \mathbb{R}^n, y \in \mathbb{R}\}$ soporta al epígrafe de f en $(\bar{x}, f(\bar{x}))$. En particular $f(x) \geq f(\bar{x}) + \langle \xi, x - \bar{x} \rangle$ para toda $x \in S$ (ξ , es un subgradiente de f en \bar{x}).
Prueba:

Puesto que f es una función convexa, por la proposición 1.1.3.7, $\text{epí}(f)$ es un conjunto convexo. Además $(\bar{x}, f(\bar{x})) \in \partial(\text{epí}(f))$ para toda $\bar{x} \in S$. Por el teorema 1.1.2.36, existe un vector no nulo $(p, \mu) \in \mathbb{R}^n \times \mathbb{R}$, tal que $\langle p, x - \bar{x} \rangle + \mu(y - f(\bar{x})) \leq 0$ para toda $(x, y) \in \text{epí}(f)$. Observe que $\mu \leq 0$, ya que en caso contrario, se podría escoger y suficientemente grande, de tal manera que la desigualdad establecida conduzca a una contradicción. También si $\mu = 0$, entonces $\langle p, x - \bar{x} \rangle \leq 0$ para todo $x \in S$ y puesto que $\bar{x} \in \text{Int}(S)$, existe $\lambda > 0$, tal que $\bar{x} + \lambda p \in S$ y por lo tanto $\langle p, \bar{x} + \lambda p - \bar{x} \rangle = \langle p, \lambda p \rangle = \lambda \|p\|^2 \leq 0$, de donde $p = 0$ y así $(p, \mu) = 0$, lo cual contradice el hecho de que $(p, \mu) \neq 0$. Sea $\xi = \frac{p}{|\mu|}$, entonces $\langle \xi, x - \bar{x} \rangle - y + f(\bar{x}) \leq 0$

para toda $(x, y) \in \text{epí}(f)$. En particular, el hiperplano $Y=f(x) + \langle \xi, x-\bar{x} \rangle$ soporta a $\text{epí}(f)$ en el punto $(\bar{x}, f(\bar{x}))$. Si $y=f(x)$, entonces $f(x) \geq f(\bar{x}) + \langle \xi, x-\bar{x} \rangle$ para todo $x \in S$.

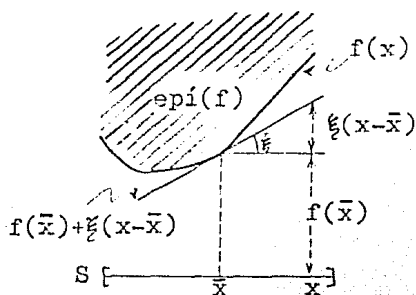


Figura 21. Interpretación geométrica del teorema 1.1.3.9

En la figura anterior se observa que un subgradiente es la pendiente de una recta (un hiperplano) de soporte del epígrafo de la función f en el punto $\bar{x} \in S$.

Un concepto más general de diferenciabilidad que el de derivada de Gateaux, es el concepto de derivada de Fréchet.

Definición. 1.1.3.10 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$, abierto y sea $f: S \rightarrow \mathbb{R}$. Se dice que la función f es diferenciable en el punto $x \in S$ en el sentido de Fréchet, si existe una transformación lineal continua $Tf(x): \mathbb{R}^n \rightarrow \mathbb{R}$, $d \mapsto (Tf(x))(d)$ con $d \in \mathbb{R}^n$, tal que si $x+d \in S$, se tiene que $f(x+d) - f(x) = (Tf(x))(d) + o(\|d\|)$, donde $o(\|d\|)$ significa que $|o(\|d\|)|/\|d\| \rightarrow 0$ cuando $\|d\| \rightarrow 0$, $o(\|d\|)$ es llamado un infinitésimo de orden superior al primero con respecto de $\|d\|$.

Si f es Fréchet-diferenciable en cada punto de S , se dice entonces que f es Fréchet diferenciable en S .

Se puede probar que $(Tf(x))(d) = \langle \text{grad}f(x), d \rangle$, donde $\text{grad}f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$. Note que para que una función sea Fréchet-diferenciable no basta con que $\text{grad}f(x)$ exista; sino que es necesario que las derivadas parciales sean continuas.

Definición. 1.1.3.11 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$, abierto y sea $f: S \rightarrow \mathbb{R}$. Se dice que la función f es dos veces Fréchet-diferenciable en el punto $x \in S$, si existe una transformación lineal continua $Tf(x): \mathbb{R}^n \rightarrow \mathbb{R}$, $d \in \mathbb{R}^n \mapsto (Tf(x))(d)$ y una transformación bilineal continua $Bf(x): \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(d, d) \mapsto (Bf(x))(d, d)$ tal que si $x+d \in S$, entonces $f(x+d) - f(x) = (Tf(x))(d) + (Bf(x))(d, d) + o(\|d\|^2)$
 $= \langle \text{grad}f(x), d \rangle + (Bf(x))(d, d) + o(\|d\|^2)$;
 donde $o(\|d\|^2)$ satisface $|o(\|d\|^2)|/\|d\| \rightarrow 0$ cuando $\|d\| \rightarrow 0$, $o(\|d\|^2)$ es llamado un infinitésimo de orden superior al segundo con respecto de $\|d\|$.

Si una función f es Fréchet-diferenciable en cada punto de S , se dice entonces que f es Fréchet-diferenciable en S .

Se puede probar también que $(Bf(x))(d, d) = \frac{1}{2} \langle d, H(x)d \rangle$, donde $H(x)$ es la matriz Hessiana de f en x ; $H(x) = \left[\frac{\partial^2 f}{\partial x_k \partial x_j} \right]_{1 \leq k, j \leq n}$

Las funciones Fréchet-diferenciables definidas sobre conjuntos convexos poseen las siguientes propiedades enunciadas en dos proposiciones, las pruebas se omiten y pueden ser encontradas en la referencia .

Teorema. 1.1.3.12 (Teorema del valor medio)

Sea S un subconjunto convexo, abierto y no vacío de \mathbb{R}^n y sea $f: S \rightarrow \mathbb{R}$ una función Fréchet-diferenciable en S . Entonces para todo $x, y \in S$, se tiene $f(y) = f(x) + \langle \text{grad}f(z), y-x \rangle$, donde $z = \lambda x + (1-\lambda)y$ para algún $\lambda \in (0, 1)$.

Teorema. 1.1.3.13 (Teorema de Taylor)

Sea S un subconjunto convexo, abierto y no vacío de \mathbb{R}^n y sea

$f: S \rightarrow \mathbb{R}$ una función dos veces Fréchet-diferenciable, entonces para todo $x, y \in S$, se tiene $f(y) = f(x) + \langle \text{grad}f(x), y-x \rangle + \frac{1}{2} \langle y-x, H(z)(y-x) \rangle$, donde $H(z)$ es la matriz Hessiana de f en z y $z = \lambda x + (1-\lambda)y$ para algún $\lambda \in (0,1)$.

A continuación, se presenta un resultado que caracteriza a las funciones convexas en términos de la matriz Hessiana.

Lema. 1.1.3.14 Sea $S \subset \mathbb{R}^n$, $S \neq \emptyset$, abierto y convexo y sea $f: S \rightarrow \mathbb{R}$ una función Fréchet-diferenciable en el punto $\bar{x} \in S$. Entonces el conjunto de subgradiientes de f en \bar{x} , es el conjunto que consiste sólo del elemento $\text{grad}f(\bar{x})$.

Prueba:

Por el teorema 1.1.3.9, el conjunto de subgradiientes de f en \bar{x} es no vacío. En consecuencia, existe $\xi \in \mathbb{R}^n$, tal que para $\lambda > 0$ suficientemente pequeña $f(\bar{x} + \lambda d) \geq f(\bar{x}) + \lambda \langle \xi, d \rangle$, donde $d \in \mathbb{R}^n$ es un vector arbitrario. También por definición de Fréchet-diferenciabilidad de f en \bar{x} , se tiene que $f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \langle \text{grad}f(\bar{x}), d \rangle + o(\lambda)$. Después de restar las dos expresiones para $f(\bar{x} + \lambda d)$, dividir entre $\lambda > 0$ y tomar el límite cuando $\lambda \rightarrow 0$, se sigue que $\langle \xi - \text{grad}f(\bar{x}), d \rangle \leq 0$, escogiendo $d = \xi - \text{grad}f(\bar{x})$, se tiene que $\| \xi - \text{grad}f(\bar{x}) \|^2 \leq 0$, de donde $\xi = \text{grad}f(\bar{x})$. Esto completa la prueba.

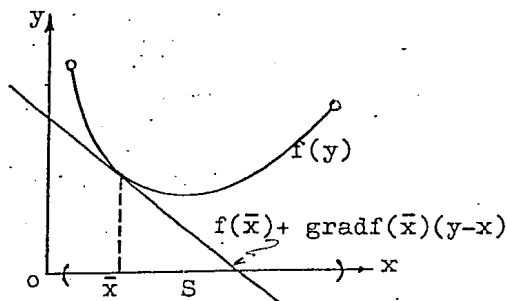


Figura 22. Interpretación geométrica del lema 1.1.3.9

Si $S \subset \mathbb{R}^n$, $S \neq \emptyset$ es un conjunto abierto y convexo, y si $f: S \rightarrow \mathbb{R}$ es una función convexa Fréchet-diferenciable, entonces por el teorema 1.1.3.9 y por el lema anterior, se tiene que $f(x) \geq f(\bar{x}) + \langle \text{grad}f(\bar{x}), x - \bar{x} \rangle$ para toda $x, \bar{x} \in S$.

Teorema. 1.1.3.15 Sea S un subconjunto no vacío, abierto y convexo de \mathbb{R}^n y $f: S \rightarrow \mathbb{R}$ una función convexa dos veces Fréchet-diferenciable sobre S . Entonces f es una función convexa si, y sólo si la matriz Hessiana $H(\bar{x})$ es semidefinida positiva en cada punto $\bar{x} \in S$, es decir, $\langle y, H(\bar{x})y \rangle \geq 0$ para todo $y \in \mathbb{R}^n$ y para todo $\bar{x} \in S$.

Prueba:

Sea $\bar{x} \in S$. Puesto que S es un conjunto abierto, entonces para cualquier $y \in \mathbb{R}^n$, se tiene que $\bar{x} + \lambda y \in S$ para λ suficientemente pequeña. Ahora bien, como f es convexa y Fréchet-diferenciable, entonces $f(\bar{x} + \lambda y) \geq f(\bar{x}) + \lambda \langle \text{grad}f(\bar{x}), y \rangle$ y por ser f dos veces Fréchet-diferenciable $f(\bar{x} + \lambda y) = f(\bar{x}) + \lambda \langle \text{grad}f(\bar{x}), y \rangle + \frac{1}{2} \lambda^2 \langle y, H(\bar{x})y \rangle + o(\lambda^2)$, después de restar adecuadamente las dos expresiones para $f(\bar{x} + \lambda y)$, se tiene $\frac{1}{2} \lambda^2 \langle y, H(\bar{x})y \rangle + o(\lambda^2) \geq 0$, dividiendo entre $\lambda^2 > 0$ y tomando el límite cuando $\lambda \rightarrow 0$, se obtiene el resultado pedido.

Recíprocamente, sean $x, \bar{x} \in S$ y suponga que $H(x)$ y $H(\bar{x})$ son matrices semidefinidas positivas. Por el teorema del valor medio se tiene que $f(x) = f(\bar{x}) + \langle \text{grad}f(\bar{x}), x - \bar{x} \rangle + \frac{1}{2} \langle x - \bar{x}, H(\hat{x})(x - \bar{x}) \rangle$, donde $\hat{x} = \lambda \bar{x} + (1 - \lambda)x$ para algún $\lambda \in (0, 1)$ y como $\hat{x} \in S$, se tiene que $f(x) \geq f(\bar{x}) + \langle \text{grad}f(\bar{x}), x - \bar{x} \rangle$. Sea $y \in S$ cualquiera, entonces también es válido que $f(y) \geq f(\bar{x}) + \langle \text{grad}f(\bar{x}), y - \bar{x} \rangle$, multiplicando la desigualdad para $f(x)$ por cualquier $\alpha \in (0, 1)$ y la de $f(y)$ por $1 - \alpha$, y sumándolas, se obtiene $\alpha f(x) + (1 - \alpha)f(y) \geq f(\bar{x}) + \langle \text{grad}f(\bar{x}), \alpha x + (1 - \alpha)y - \bar{x} \rangle$, si ahora se toma $\bar{x} = \alpha x + (1 - \alpha)y$, se tiene $\alpha f(x) + (1 - \alpha)f(y) \geq f(\alpha x + (1 - \alpha)y)$ para cualquier $\alpha \in (0, 1)$.

A continuación se estudian condiciones necesarias y suficientes de soluciones óptimas de problemas de programación no lineal convexa:

Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y S un conjunto convexo no vacío de \mathbb{R}^n . Considere el problema

(I). Minimizar $f(x)$
sujeto a $x \in S$

Se recuerda aquí que $x^* \in S$ es un mínimo local del problema (I), si existe $\epsilon > 0$, tal que $f(x^*) \leq f(x)$ para todo $x \in S$ que satisficiera $\|x - x^*\| < \epsilon$. También se dice que $x^* \in S$ es un mínimo global, si $f(x^*) \leq f(x)$ para toda $x \in S$.

Teorema. 1.1.3.16 Considere el problema (I). Entonces $x^* \in S$ es una solución óptima global de (I) si, y sólo si la función f tiene un subgradiente ξ en x^* tal que $\langle \xi, x - x^* \rangle \geq 0$ para toda $x \in S$.
Prueba:

(Suficiencia) Sea ξ un subgradiente de f en x^* , entonces $f(x) \geq f(x^*) + \langle \xi, x - x^* \rangle$ para todo $x \in S$. Pero por hipótesis $\langle \xi, x - x^* \rangle \geq 0$ para toda $x \in S$, de donde $f(x) \geq f(x^*) + \langle \xi, x - x^* \rangle \geq f(x^*)$ para toda $x \in S$, lo cual implica que x^* es un óptimo global de (I).

(Necesidad) Suponga que x^* es solución óptima global de (I) y considere los dos conjuntos siguientes: $Q_1 = \{(x - x^*, y); x \in \mathbb{R}^n, y > f(x) - f(x^*)\}$ y $Q_2 = \{(x - x^*, y); x \in S, y \leq 0\}$. Los conjuntos Q_1 y Q_2 son conjuntos convexos, ajenos y no vacíos. Por el teorema 1.1.2.37, existe un hiperplano que los separa, es decir, existen un vector $(p, \mu) \in \mathbb{R}^{n+1}$, $(p, \mu) \neq 0$ y un escalar $\alpha \in \mathbb{R}$, tales que:

- (i). $\langle p, x - x^* \rangle + \mu y \leq \alpha$ para todo $(x - x^*, y) \in Q_1$,
- (ii). $\langle p, x - x^* \rangle + \mu y \geq \alpha$ para todo $(x - x^*, y) \in Q_2$.

Ahora bien, note que el vector $(0, 0) \in Q_2$, ya que $x^* \in S$ y $y' = 0$ satisfacen los requisitos de pertenencia. Por lo tanto de (ii) se obtiene que $\alpha \leq 0$. Por otra parte, $S \subset \mathbb{R}^n$ y por lo tanto $(0, y) \in Q_1$ implica $\mu y' \leq \alpha$ para todo $(0, y) \in Q_1$. Pero al mismo tiempo $(0, y) \in Q_1$ permite que $y > f(x) - f(x^*) \geq 0$, ya que, por hipótesis $f(x^*) \leq f(x)$ cada vez que $x \in S$. Note que la desigualdad $\mu y' \leq \alpha$ tiene que mantenerse para cualquier $y' > 0$, lo cual sólo puede cumplirse

si $\mu \leq 0$ y $\alpha \geq 0$, de donde $\alpha \leq 0 \leq \alpha$, es decir $\alpha = 0$. Se verá ahora que $\mu < 0$. Suponga que $\mu = 0$, entonces de (i) se tiene $\langle p, x - x^* \rangle \leq 0$ para toda $x \in \mathbb{R}^n$, si se toma $x = x^* + p$, se tendría $\langle p, x - x^* \rangle = \langle p, x^* + p - x^* \rangle = \|p\|^2 \leq 0$, lo cual implica que $p = 0$ contradiciendo el hecho de que $(p, \mu) \neq 0$. Por lo tanto $\mu < 0$ y dividiendo (ii) entre $-\mu$ se obtiene que $-\langle \frac{p}{\mu}, x - x^* \rangle - \gamma \geq 0$ para toda $(x - x^*, \gamma) \in Q_2$, es decir $-\langle \frac{p}{\mu}, x - x^* \rangle - \gamma \geq 0$ para toda $x \in S$ y $\gamma \leq 0$. Si se hace $\gamma = 0$ y $\xi = -\frac{1}{\mu} p$, se obtiene que $\langle \xi, x - x^* \rangle \geq 0$ para toda $x \in S$. Finalmente, se verá que ξ es un subgradiente de f en x^* , dividiendo (i) entre $-\mu$ se obtiene que $\langle \xi, x - x^* \rangle - \gamma \leq 0$ para toda $x \in \mathbb{R}^n$, tal que $\gamma > f(x) - f(x^*)$. Sea $\lambda > 0$, entonces $\gamma = f(x) - f(x^*) + \lambda > f(x) - f(x^*)$ para toda $x \in \mathbb{R}^n$ y por lo tanto $\langle \xi, x - x^* \rangle - f(x) + f(x^*) - \lambda \leq 0$ para todo $x \in \mathbb{R}^n$ y $\lambda > 0$, cuando $\lambda \rightarrow 0$ se tiene la conclusión pedida.

Corolario. 1.1.3.17 Bajo las hipótesis del teorema anterior, si en adición el conjunto S es abierto. Entonces $x^* \in S$ es una solución óptima del problema (I) si, y sólo si existe un subgradiente $\xi = 0$ de f en x^* .

Prueba:

Por el teorema anterior, $x^* \in S$ es una solución óptima si, y sólo si f tiene un subgradiente ξ en x^* , tal que $\langle \xi, x - x^* \rangle \geq 0$ para todo $x \in S$. Como S es abierto, $x = x^* + \lambda \xi \in S$ con $\lambda > 0$ suficientemente pequeño, entonces $\langle \xi, x - x^* \rangle = \lambda \langle \xi, \xi \rangle = \lambda \|\xi\|^2 \geq 0$, de donde $\xi = 0$.

Corolario. 1.1.3.18 Además de las hipótesis del teorema 1.1.3.16, suponga que S es abierto y f es Fréchet-diferenciable. Entonces $x^* \in S$ es una solución óptima global si, y sólo si $\text{grad}f(x^*) = 0$.

Las siguientes dos proposiciones son útiles en la identificación de soluciones óptimas globales únicas del problema (I).

Proposición. 1.1.3.19 Sea $x^* \in S$ una solución óptima local del problema (I), entonces $x^* \in S$ es una solución óptima global.

Prueba:

Suponga que por el contrario, que $x^* \in S$ no es una solución óptima global del problema (I), entonces existe algún $\bar{x} \in S$, tal que $f(\bar{x}) < f(x^*)$. Ahora bien, si x^* es solución óptima local de (I), entonces existe $\epsilon > 0$, tal que para toda $x \in S$ que satisface $\|x - x^*\| \leq \epsilon$ se cumple que $f(x^*) \leq f(x)$. Sea $\lambda \in (0, 1)$ suficientemente pequeña, tal que $\lambda \|\bar{x} - x^*\| < \frac{\epsilon}{2}$, entonces $\lambda \bar{x} + (1 - \lambda)x^* \in S$ y satisface $\|\lambda \bar{x} + (1 - \lambda)x^* - x^*\| = \lambda \|\bar{x} - x^*\| < \frac{\epsilon}{2}$ y en consecuencia $f(\lambda \bar{x} + (1 - \lambda)x^*) \leq \lambda f(\bar{x}) + (1 - \lambda)f(x^*) < f(x^*)$, lo cual contradice el hecho de que $x^* \in S$ sea una solución óptima local de (I).

Proposición. 1.1.3.20 Considere el problema (I), si en adición f es estrictamente convexa y si x^* es una solución óptima local, entonces $x^* \in S$ es una solución óptima global única. prueba:

Por la proposición anterior, se sigue que $x^* \in S$ es una solución óptima global. Para ver la unicidad suponga que por el contrario que existe $\bar{x} \in S$ $\bar{x} \neq x^*$, tal que $f(\bar{x}) = f(x^*)$. Sea $\lambda \in (0, 1)$ fijo, entonces por la convexidad de S , $\lambda \bar{x} + (1 - \lambda)x^* \in S$ y por ser f una función estrictamente convexa se tiene que $f(\lambda \bar{x} + (1 - \lambda)x^*) < \lambda f(\bar{x}) + (1 - \lambda)f(x^*) = f(x^*)$, con lo cual $\lambda \bar{x} + (1 - \lambda)x^* \in S$ contradice la optimalidad global de x^* .

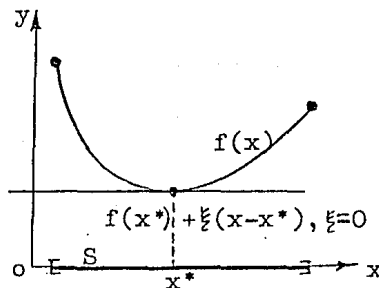


Figura 23. Solución óptima global única de una función estrictamente convexa.

Hasta este momento se han estudiado funciones y conjuntos convexos y se ha visto, que éstos poseen propiedades muy importantes dentro de la teoría de optimización. Sin embargo, como ya se había mencionado, en la práctica es imposible esperar que todas las funciones sean convexas. Lo que interesa es pues, investigar hasta donde se pueden relajar los supuestos de convexidad manteniendo las propiedades deseables para efectos de óptimalidad.

Definición. 1.1.3.21 Sea $f:S \rightarrow \mathbb{R}$ una función definida sobre un conjunto convexo no vacío SCR^n . Se dice que la función f es cuasi-convexa si, para toda $x, y \in S$ se satisface que $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$ para cualquier $\lambda \in [0, 1]$. Una función f es llamada cuasi-cóncava si $-f$ es cuasi-convexa.

La definición anterior dice que, la función f es cuasi-convexa si $f(y) \geq f(x)$ implica que $f(y) \geq f(\lambda x + (1-\lambda)y)$ para todo $\lambda \in [0, 1]$.

Una de las caracterizaciones más importantes de las funciones cuasi-convexas viene dada por el siguiente teorema:

Teorema. 1.1.3.22 Sea $f:S \rightarrow \mathbb{R}$ una función definida sobre un subconjunto convexo no vacío S de \mathbb{R}^n . Entonces f es cuasi-convexa si, y sólo si el conjunto $Q_\alpha = \{x \in S; f(x) \leq \alpha\}$ es un conjunto convexo para cualquier número real α .

Prueba:

Suponga que la función f es cuasi-convexa y sean $x, y \in Q_\alpha$, $\alpha \in \mathbb{R}$. Como $x, y \in Q_\alpha$ se tiene que $x, y \in S$ y que $\max\{f(x), f(y)\} \leq \alpha$. Sea $z = \lambda x + (1-\lambda)y$, $\lambda \in [0, 1]$. Entonces $z \in S$, ya que, el conjunto S es convexo, y como f es cuasi-convexa, se tiene que $f(z) \leq \max\{f(x), f(y)\} \leq \alpha$, es decir $z \in Q_\alpha$. Esto implica que Q_α es un conjunto convexo para α un número real arbitrario.

Recíprocamente, suponga que Q_α es un conjunto convexo para cualquier $\alpha \in \mathbb{R}$. Sean $x, y \in S$, $\lambda \in [0, 1]$ y $z = \lambda x + (1-\lambda)y$. Si $\alpha = \max\{f(x), f(y)\}$, se tiene entonces que $x, y \in Q_\alpha$ y como por hipótesis

Q_α es un conjunto convexo, $f(z) \leq \alpha = \max\{f(x), f(y)\}$. Por lo tanto f es cuasi-convexa.

A continuación se proporciona la definición de función estrictamente cuasi-convexa.

Definición. 1.1.3.23 Sea $f: S \rightarrow \mathbb{R}$ una función definida sobre un conjunto convexo no vacío S de \mathbb{R}^n . Se dice que f es estrictamente cuasi-convexa si para cualesquier $x, y \in S$ con $f(x) \neq f(y)$, se tiene que $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$ para todo $\lambda \in (0, 1)$. Una función f es llamada estrictamente cuasi-cóncava si $-f$ es estrictamente cuasi-convexa.

Observe que cualquier función convexa es una función estrictamente cuasi-convexa. En efecto, sean $x, y \in S$ con $f(x) \neq f(y)$ y suponga que en particular $f(y) > f(x)$, si f es convexa, entonces $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) < f(y) = \max\{f(x), f(y)\}$ con $\lambda \in (0, 1)$.

La diferencia principal entre una función estrictamente cuasi-convexa y una que es sólomente cuasi-convexa, es que en la primera no se admiten puntos que no sean mínimos globales en los que el gradiente desaparece. Este hecho se demuestra en el siguiente teorema.

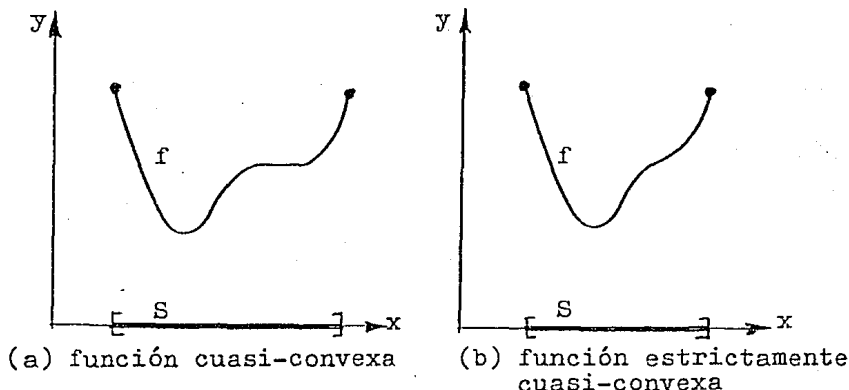


Figura 24. Generalizaciones de funciones convexas.

Teorema 1.1.3.24 Considere el problema (I). Si en adición f es una función estrictamente cuasi-convexa. Entonces si $x^* \in S$ es una solución óptima local del problema (I), también es una solución óptima global de (I).

Prueba:

Suponga que por el contrario, que existe $\bar{x} \in S$, tal que $f(\bar{x}) < f(x^*)$, como S es un conjunto convexo, se tiene que $z = \lambda \bar{x} + (1-\lambda)x^* \in S$ para toda $\lambda \in (0,1)$. Ahora bien, como $x^* \in S$ es solución óptima local, existe $0 < \epsilon < 1$ tal que $f(x^*) \leq f(\lambda \bar{x} + (1-\lambda)x^*)$ para toda $\lambda \in (0, \epsilon)$. Pero como f es estrictamente cuasi-convexa y $f(\bar{x}) < f(x^*)$ se tiene que también $f(\lambda \bar{x} + (1-\lambda)x^*) < f(x^*)$, lo cual conduce a la contradicción de que $f(x^*) < f(x^*)$. Por lo tanto x^* es una solución óptima global del problema (I).

Un hecho curioso es que, contrario a lo que pudiera suponerse, una función estrictamente cuasi-convexa no es necesariamente cuasi-convexa. Para ver esto, examine la función siguiente:

$$f(x) = \begin{cases} 1, & \text{si } x=0 \\ 0, & \text{si } x \neq 0 \end{cases}$$

De acuerdo a la definición dada, la función f es estrictamente cuasi-convexa. Sin embargo, si $x=1$ y $y=-1$, se tiene que $f(\frac{1}{2}x + \frac{1}{2}y) = f(0) = 1 > \max\{f(x), f(y)\} = 0$, lo cual se debe a que $f(x) = f(y)$ y $x \neq y$.

Definición. 1.1.3.25 Sea $f: S \rightarrow \mathbb{R}$ una función definida sobre un subconjunto convexo no vacío S de \mathbb{R}^n . Se dice que f es fuertemente cuasi-convexa, si para todo $x, y \in S$ con $x \neq y$ se verifica que: $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$ para toda $\lambda \in (0,1)$.

Se dice que f es fuertemente cuasi-cóncava si $-f$ es fuertemente cuasi-convexa.

Proposición. 1.1.3.26

(i). Toda función estrictamente convexa es fuertemente cuasi-convexa.

(ii). Toda función fuertemente cuasi-convexa es estrictamente cuasi-convexa.

En la figura siguiente se muestra una función que es estrictamente cuasi-convexa, pero no fuertemente cuasi-convexa.

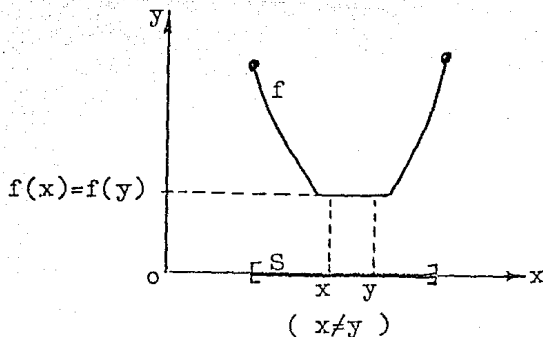


Figura 25. Ejemplo de una función que es estrictamente cuasi-convexa pero no fuertemente cuasi-convexa.

La propiedad más importante de las funciones fuertemente cuasi-convexas se establece en el siguiente teorema:

Teorema. 1.1.3.27 Considere el problema (I), si en adición la función f es fuertemente cuasi-convexa. Si $x^* \in S$ es una solución óptima local del problema (I), entonces $x^* \in S$ es una solución óptima global única.

Prueba:

Por la proposición anterior, si f es fuertemente cuasi-convexa, entonces f es estrictamente cuasi-convexa. Por el teorema 1.1.3.24, si x^* es una solución óptima local del problema (I), entonces x^* es una solución óptima global de (I).

Para ver la unicidad, suponga que existe $\bar{x} \in S$, $\bar{x} \neq x^*$, tal que

$f(\bar{x})=f(x^*)$, entonces como f es una función fuertemente cuasi-convexa, $f(\lambda\bar{x}+(1-\lambda)x^*) < \max\{f(\bar{x}), f(x^*)\}=f(x^*)$ para toda $\lambda \in (0, 1)$, lo cual contradice el hecho de que $x^* \in S$ sea una solución óptima global de (I).

Por último, en esta sección se estudian las funciones pseudo-convexas.

Definición. 1.1.3.28 Sea SCR^n un conjunto convexo, abierto y no vacío, y sea $f: S \rightarrow \mathbb{R}$ una función Fréchet-diferenciable sobre el conjunto S . Se dice que la función f es pseudo-convexa, si para todo $x, y \in S$, $\langle \text{grad}f(x), y-x \rangle \geq 0$ implica que $f(y) \geq f(x)$. Equivalentemente, si $f(y) < f(x)$ implica $\langle \text{grad}f(x), y-x \rangle < 0$.

Se dice que una función f es estrictamente pseudo-convexa si para todo $x, y \in S$ con $x \neq y$, $\langle \text{grad}f(x), y-x \rangle \geq 0$ implica que $f(y) > f(x)$, esto equivale a decir que $f(y) \leq f(x)$ implica que $\langle \text{grad}f(x), y-x \rangle < 0$.

Para las respectivas definiciones de funciones pseudo-cóncavas basta cambiar los sentidos de las desigualdades.

Proposición. 1.1.3.29 Considere el problema (I) y suponga además que la función f es pseudo-convexa, entonces $\text{grad}f(x^*)=0$ es condición necesaria y suficiente para que $x^* \in S$ sea una solución óptima global.

Prueba:

La necesidad es inmediata a partir del corolario 1.1.3.18.

Para ver suficiencia, suponga que $\text{grad}f(x^*)=0$, entonces $\langle \text{grad}f(x^*), x-x^* \rangle = 0$ para todo $x \in S$, lo cual por definición implica que $f(x^*) \leq f(x)$ para toda $x \in S$.

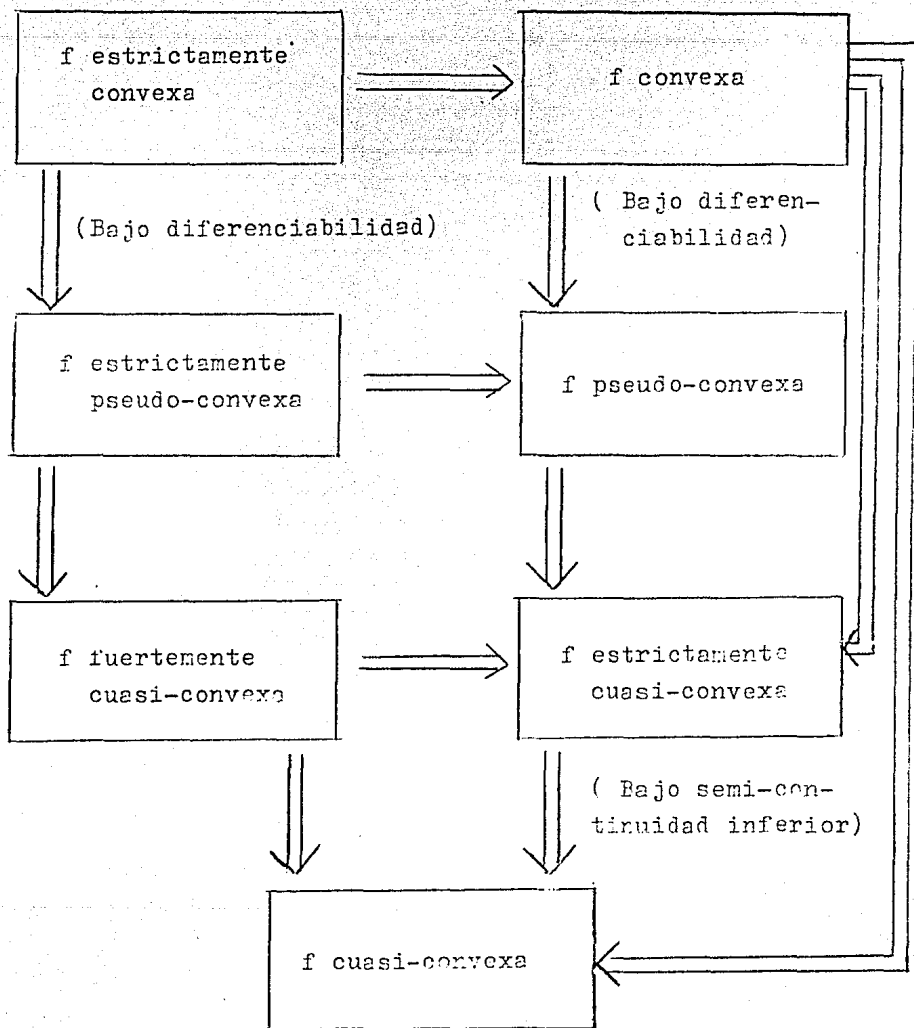


Figura 26. Relación entre los diferentes tipos de convexidad.

1.2 CONDICIONES DE OPTIMALIDAD

1.2.1 INTRODUCCION

La obtención de condiciones suficientes para soluciones óptimas locales o globales de problemas de programación no lineal, es indispensable en el desarrollo de criterios de optimalidad para la formulación de algoritmos de solución. Asimismo, la obtención de condiciones necesarias, es útil en la caracterización de dichas soluciones óptimas.

Al principio de la década de los cuarentas, Fritz John presentó los primeros resultados generales sobre condiciones de optimalidad de problemas de programación no lineal con restricciones que pueden ser igualdades y/o desigualdades. Poco después, en Princeton John Von Neumann y A.W. Tucker junto con dos de sus discípulos H. Kuhn y D. Gale, introduciendo la noción de regularidad en las restricciones obtuvieron resultados, que si no son más generales que los de Fritz John, son mas útiles en la práctica, además de que tuvieron mayor divulgación. Las condiciones de Fritz John pueden satisfacerse trivialmente en puntos no óptimos cuando el gradiente de la función objetivo y/o los gradientes de las funciones que definen las restricciones se anulan. Además de que los resultados obtenidos por H. Kuhn y A.W. Tucker esclarecen el importante concepto de dualidad lagrangeana originalmente propuesto por Wolfe.

1.2.2 CONDICIONES DE FRITZ JOHN Y KUHN-TUCKER

En la sección anterior fueron estudiadas condiciones de optimalidad para el problema:

(I). Minimizar $f(x)$
 sujeto a $x \in S$

donde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa diferenciable y S es un subconjunto abierto, convexo y no vacío de \mathbb{R}^n , y se vió que una condición necesaria y suficiente para que un punto $x^* \in S$ fuera solución óptima global del problema planteado, es que $\langle \text{grad}f(x^*), x - x^* \rangle \geq 0$ para toda $x \in S$.

En esta sección, la naturaleza del conjunto S será especificada explícitamente. En particular, se considerarán problemas con restricciones que son igualdades y/o desigualdades. Las condiciones necesarias de optimalidad serán obtenidas sin hipótesis de convexidad. No obstante, se verá que bajo condiciones relajadas de convexidad, las condiciones necesarias establecidas serán también suficientes.

A continuación se presenta el concepto de cono de direcciones factibles, el cual es útil en la obtención de condiciones geométricas necesarias de optimalidad.

Definición. 1.2.2.1

(i). Sea S un subconjunto no vacío de \mathbb{R}^n y sea $\bar{x} \in S$, se dice que un vector $d \neq 0$ es una dirección factible en \bar{x} , si existe $\delta > 0$, tal que $x = \bar{x} + \lambda d \in S$ para todo $\lambda \in (0, \delta)$.

(ii). El conjunto $D(\bar{x})$ de todas las direcciones factibles en \bar{x} , se le conoce como el cono de direcciones factibles de S en \bar{x} ; es decir, $D(\bar{x}) = \{d \neq 0; \text{ existe } \delta > 0 \text{ tal que } \bar{x} + \lambda d \in S \text{ para toda } \lambda \in (0, \delta)\}$.

De la definición anterior, es claro que, una dirección $d \in \mathbb{R}^n$, $d \neq 0$ es factible, a partir de $\bar{x} \in S$, si es posible desplazarse en esa dirección y obtener otros puntos pertenecientes al conjunto S .

Proposición. 1.2.2.2 Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable en el punto \bar{x} . Si existe un vector $d \in \mathbb{R}^n$, $d \neq 0$, tal que $\langle \text{grad}f(\bar{x}), d \rangle < 0$. Entonces existe $\delta > 0$ tal que $f(\bar{x} + \lambda d) < f(\bar{x})$ para todo $\lambda \in (0, \delta)$ (El vector $d \in \mathbb{R}^n$, es llamado una dirección factible de descenso de f en \bar{x}).

Prueba:

Si la función f es diferenciable en el punto \bar{x} , entonces por el teorema de Taylor, $f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \langle \text{grad}f(\bar{x}), d \rangle + o(\lambda)$, donde $o(\lambda)/\lambda \rightarrow 0$ cuando $\lambda \rightarrow 0$, así para $\lambda > 0$, $\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \langle \text{grad}f(\bar{x}), d \rangle + o(\lambda)/\lambda$ y puesto que por hipótesis $\langle \text{grad}f(\bar{x}), d \rangle < 0$ y $o(\lambda)/\lambda \rightarrow 0$ cuando $\lambda \rightarrow 0$, se tiene que existe un $\delta > 0$, tal que $\langle \text{grad}f(\bar{x}), d \rangle + o(\lambda)/\lambda < 0$ para toda $\lambda \in (0, \delta)$, de donde se sigue el resultado pedido.

Si se considera ahora el problema

(I'). Minimizar $f(x)$
sujeto a $x \in S$,

donde S es un subconjunto no vacío de \mathbb{R}^n y $f: S \rightarrow \mathbb{R}$ una función. Se sabe que si S es un conjunto abierto y f es diferenciable en el punto $\bar{x} \in S$, y $d \in D(\bar{x})$ es tal que $\langle \text{grad}f(\bar{x}), d \rangle < 0$, entonces por la proposición 1.2.2.2, se tiene que es posible desplazarse en la dirección d y obtener un punto $x \in S$, tal que $f(x) < f(\bar{x})$. De este hecho se espera que, para que $\bar{x} \in S$ sea una solución óptima local del problema (I') con las hipótesis adicionalmente hechas, es necesario que $\langle \text{grad}f(\bar{x}), d \rangle \geq 0$ para toda $d \in D(\bar{x})$. Es decir, $\langle \text{grad}f(\bar{x}), d \rangle < 0$ implica que $d \notin D(\bar{x})$. Esta idea se precisa en el siguiente teorema.

Teorema. 1.2.2.3 Considere el problema (I') y suponga además que S es un conjunto abierto y que la función f es diferenciable en el punto $x^* \in S$. Defina $F(x^*) = \{d \neq 0; \langle \text{grad} f(x^*), d \rangle < 0\}$. Si x^* es una solución óptima local del problema (I'), entonces $F(x^*) \cap D(x^*) = \emptyset$.

Prueba:

Suponga lo contrario, es decir, existe un $d \in F(x^*) \cap D(x^*)$, entonces por la proposición 1.2.2.2, existe un $\eta > 0$, tal que $f(x^* + \lambda d) < f(x^*)$ para toda $\lambda \in (0, \eta)$ y por la definición de dirección factible, existe $\delta > 0$, tal que $x^* + \lambda d \in S$ para toda $\lambda \in (0, \delta)$. Tome $\epsilon = \min\{\eta, \delta\} > 0$, se tiene entonces que $f(x^* + \lambda d) < f(x^*)$ y $x^* + \lambda d \in S$ para toda $\lambda \in (0, \epsilon)$, lo cual contradice el hecho de que $x^* \in S$ sea una solución óptima local del problema (I') con las hipótesis adicionales.

Ahora suponga que el conjunto S , se especifica de la manera siguiente:

$$S = \{x \in X; g_i(x) \leq 0, i=1, 2, \dots, m\} \subset \mathbb{R}^n,$$

donde $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ para $i=1, 2, \dots, m$ y X es un conjunto abierto y no vacío de \mathbb{R}^n . Considere ahora, el problema de programación no lineal con restricciones definidas por un conjunto de desigualdades:

(II). Minimizar $f(x)$

$$\text{sujeto a } g_i(x) \leq 0 \quad i=1, 2, \dots, m \\ x \in X$$

Las condiciones necesarias de optimalidad cuando el conjunto S se especifica con estas desigualdades se proporcionan en el siguiente teorema:

Teorema. 1.2.2.4 Considere el problema (II). Sea $x^* \in X$ una solución factible y sea $I = \{i; g_i(x^*) = 0\}$. Suponga que las funciones $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ para $i \in I$ son continuas en el punto $x^* \in X$ y suponga

ga además que las funciones g_i son diferenciables en $x^* \in X$ para $i \in I$ y que f es una función diferenciable en x^* . Sean $F(x^*) = \{d \neq 0; \langle \text{grad}f(x^*), d \rangle < 0\}$ y $G(x^*) = \{d \neq 0; \langle \text{grad}g_i(x^*), d \rangle < 0 \text{ para toda } i \in I\}$. Si x^* es una solución óptima local del problema (II), entonces $F(x^*) \cap G(x^*) = \emptyset$.

Prueba:

Sea $d \in G(x^*)$, como $x^* \in X$ y X es un conjunto abierto de \mathbb{R}^n , existe $\epsilon > 0$, tal que $x^* + \lambda d \in X$ para toda $\lambda \in (0, \epsilon)$. Además, para $i \in I$, por continuidad de las g_i , se tiene que $g_i(x^*) < 0$ y por lo tanto existe $\gamma > 0$, tal que $g_i(x^* + \lambda d) < 0$ para todo $\lambda \in (0, \gamma)$ para toda $i \in I$. Finalmente, como $d \in G(x^*)$, se tiene que $\langle \text{grad}g_i(x^*), d \rangle < 0$ y por lo tanto existe $\eta > 0$, tal que $g_i(x^* + \lambda d) < g_i(x^*) \leq 0$ para todo $\lambda \in (0, \eta)$ y para todo $i \in I$. Así, para $\delta = \min\{\epsilon, \gamma, \eta\}$, se cumple: $g_i(x^* + \lambda d) < 0$, $x^* + \lambda d \in X$ para toda $\lambda \in (0, \delta)$ y para toda $i \in \{1, 2, \dots, m\}$, lo cual indica que $d \in D(x^*)$. Pero por el teorema anterior $F(x^*) \cap D(x^*) = \emptyset$, de donde $F(x^*) \cap G(x^*) = \emptyset$.

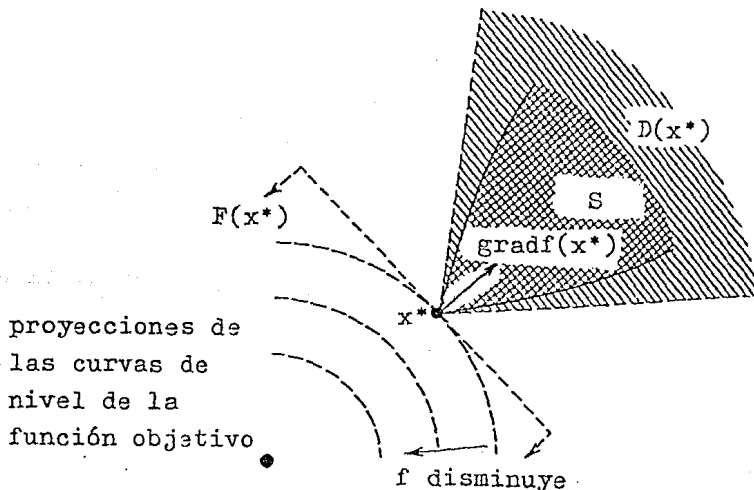


Figura 27. Ilustración de la condición necesaria geométrica $F(x^*) \cap G(x^*) = \emptyset$

En lo que sigue, se transforman las condiciones geométricas necesarias de optimalidad que se han desarrollado, a una forma más operacional. Estas condiciones son conocidas como condiciones necesarias de optimalidad de Fritz-John.

Teorema. 1.2.2.5 (Condiciones necesarias de Fritz John para el problema (II)).

Considere el problema (II) y sea x^* una solución factible, es decir $x^* \in \{x \in X; g_i(x) \leq 0, i=1,2,\dots,m\}$. Defina ahora el conjunto $I = \{i; g_i(x^*) = 0\}$ y suponga además que f y las g_i para $i \in \{1, 2, \dots, m\}$ son diferenciables en el punto x^* . Si $x^* \in X$ es una solución óptima local del problema (II). Entonces existen u_0, u_i para $i=1,2,\dots,m$, no todos cero, tales que:

$$u_0 \text{grad}f(x^*) + \sum_{i=1}^m u_i \text{grad}g_i(x^*) = 0$$

$$u_0, u_i \geq 0, i=1,2,\dots,m$$

$$u_i g_i(x^*) = 0, i=1,2,\dots,m$$

Prueba:

Como x^* es una solución óptima local sobre $S = \{x \in X; g_i(x) \leq 0, i=1,2,\dots,m\}$, por el teorema 1.2.2.4 es imposible que exista un $d \in \mathbb{R}^n$, tal que $\langle \text{grad}f(x^*), d \rangle < 0$ y $\langle \text{grad}g_i(x^*), d \rangle < 0$ para todo $i \in I$. Sea A una matriz cuyos renglones consisten de los vectores $\text{grad}f(x^*)^t$ y $\text{grad}g_i(x^*)^t$ para $i \in I$. Entonces la condición geométrica necesaria de optimalidad $F(x^*) \cap G(x^*) = \emptyset$, equivale a decir que el sistema $Ad < 0$ no tiene solución. Por el teorema de alternativa de Gordan, se tiene que existe un vector $u \geq 0, u \neq 0$, tal que $u^t A = 0$, si $u = (u_0, u_1)^t$, entonces se tiene que:

$$u^t A = u_0 \text{grad}f(x^*) + \sum_{i \in I} u_i \text{grad}g_i(x^*) = 0$$

Ahora bien, observe que para $i \in I$, $g_i(x^*) = 0$ y por lo tanto $u_i \cdot g_i(x^*) = 0$ si $i \in I$. Pero además en la conclusión del teorema sólo se requiere que algunos u_i sean diferentes de cero. Por lo tanto, haciendo $u_i = 0$ para $i \notin I$ se obtienen todas las condiciones propuestas en el enunciado del teorema.

A los escalares u_0, u_i $i=1, 2, \dots, m$, se les conoce como multiplicadores de Lagrange y a la condición $u_i g_i(x^*) = 0$ $i=1, 2, \dots, m$, como holguras complementarias.

Aunque cronológicamente las condiciones de Fritz John se obtuvieron antes que las de Kuhn-Tucker, además de que las condiciones de Fritz John son más generales, las condiciones de Kuhn-Tucker tuvieron mayor divulgación y son de hecho más útiles. Esto se debe a que es posible que se cumplan las condiciones de Fritz John trivialmente en puntos no óptimos. Shety ha construido varios de estos ejemplos. En estos casos lo que sucede es que $\text{grad}f(x^*) = 0$ y/o $\text{grad}g_i(x^*) = 0$ para alguna $i \in I$ y es posible dejar que los correspondientes multiplicadores de Lagrange sean cualquier número positivo y los demás cero, para cumplir las condiciones en puntos no óptimos.

En realidad, el único caso útil, es cuando $u_0 > 0$, ya que, si $u_0 = 0$ las condiciones de Fritz John no proporcionan información alguna sobre lo que sucede con la función objetivo en el punto x^* . Sin embargo, es posible identificar el caso en que $u_0 > 0$ si se pide cierta condición adicional sobre las restricciones. En el siguiente teorema debido a Kuhn-Tucker, que se enuncia a continuación se impone el requisito restrictional de que los vectores $\text{grad}g_i(x^*)$ con $i \in I$ sean linealmente independientes en \mathbb{R}^n para que se obtenga $u_0 > 0$.

Teorema. 1.2.2.6 (Condiciones necesarias de Kuhn-Tucker para el problema (II))

Considere el problema (II) y sea $x^* \in X$ una solución factible.

Sea $I = \{i; g_i(x^*) = 0\}$ y suponga además que las funciones f y g_i para $i \in I$ son diferenciables en el punto x^* y que las funciones g_i para $i \notin I$ son continuas en x^* . Si los vectores $\text{grad}g_i(x^*)$ con $i \in I$ son linealmente independientes en \mathbb{R}^n y si x^* es una solución óptima local del problema (II). Entonces existen escalares $u_i \geq 0$, no todos cero para $i \in I$, tales que:

$$\text{grad}f(x^*) + \sum_{i \in I} u_i \text{grad}g_i(x^*) = 0$$

$$u_i g_i(x^*) = 0 \text{ para todo } i \in I.$$

Si además, las funciones g_i son también diferenciables en el punto $x^* \in X$ cuando $i \notin I$. Se tiene que existen $u_i, i=1,2,\dots,m$, no todos cero, tales que:

$$\text{grad}f(x^*) + \sum_{i=1}^m u_i \text{grad}g_i(x^*) = 0$$

$$u_i g_i(x^*) = 0, u_i \geq 0, i=1,2,\dots,m$$

Prueba:

Si $x^* \in X$ es una solución óptima local del problema (II), las condiciones de Fritz John afirman que existen $v_0, v_i \geq 0$ no todos cero para $i \in I$, tales que:

$$v_0 \text{grad}f(x^*) + \sum_{i \in I} v_i \text{grad}g_i(x^*) = 0, \text{ y}$$

$$v_i g_i(x^*) = 0, i=1,2,\dots,m.$$

Ahora bien, como los vectores $\text{grad}g_i(x^*)$ con $i \in I$ son linealmente independientes en \mathbb{R}^n , entonces se debe tener que $v_0 > 0$, en efecto, si $v_0 = 0$, entonces los $\text{grad}g_i(x^*)$ para $i \in I$ serían linealmente dependientes. Sean

$$u_0 = \frac{v_0}{v_0} = 1 > 0, u_i = \frac{v_i}{v_0} \text{ si } i \in I, \text{ y } u_i = 0 \text{ si } i \notin I,$$

con lo cual se concluye la prueba del teorema.

Note que las condiciones de Kuhn-Tucker son un caso particular de las de Fritz John, en donde se exige un requisito restricticional para garantizar que $u_0 > 0$.

El requisito restricticional impuesto es conocido como condición de regularidad de las restricciones del problema (II).

Las condiciones necesarias de optimalidad de Kuhn-Tucker aceptan una interpretación geométrica sencilla. Se debe notar primero que cualquier vector $y = \sum_{i \in I} u_i \text{grad} g_i(x^*)$ con $u_i \geq 0$, para $i \in I$ pertenece al cono generado por los gradientes de las restricciones activas en x^* , es decir en aquellas restricciones g_i con $i \in I$ en las cuales $u_i > 0$. Entonces la condición $\text{grad} f(x^*) + \sum_{i \in I} u_i \text{grad} g_i(x^*) = 0$,

claramente equivale a decir que $-\text{grad} f(x^*)$ está en dicho cono. En consecuencia, las condiciones de Kuhn-Tucker dicen que si $x^* \in X$ es una solución óptima local del problema (II), entonces $-\text{grad} f(x^*)$ pertenece al cono generado por los vectores $\text{grad} g_i(x^*)$ para $i \in I$.

Hasta este momento, sólo se han presentado condiciones necesarias para soluciones óptimas locales del problema (II). Ahora se verá que bajo condiciones relajadas de convexidad, es posible obtener también condiciones suficientes.

Teorema. 1.2.2.7 (Condiciones suficientes de Kuhn-Tucker para el problema (II))

Considere el problema (II). Sea $x^* \in X$ una solución factible y sean $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función pseudo-convexa y $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ funciones cuasi-convexas y diferenciables en x^* para $i=1,2,\dots,m$. Si se satisfacen las condiciones de Kuhn-Tucker en $x^* \in X$. Entonces x^* es una solución óptima global del problema (II).

Prueba:

Sea $x \in S$ una solución factible cualquiera del problema (II). Entonces $g_i(x) \leq g_i(x^*) = 0$ para $i \in I$, donde $I = \{i; g_i(x^*) = 0\}$. Por la propiedad de cuasi-convexidad de las funciones g_i , se tiene que $g_i(\lambda x + (1-\lambda)x^*) = g_i(x^* + \lambda(x-x^*)) \leq \max\{g_i(x), g_i(x^*)\} = g_i(x^*) = 0$ para cualquier $\lambda \in (0,1)$ y para toda $i \in I$. Lo anterior implica que es posible desplazarse en la dirección $d = x - x^*$ sin incrementar los valores de las funciones g_i para $i \in I$.

Por la proposición 1.2.2.2, se tiene que $\langle \text{grad}g_i(x^*), d \rangle \leq 0$ para toda $i \in I$, es decir, $\langle \text{grad}g_i(x^*), x-x^* \rangle \leq 0$ para toda $i \in I$ y para toda $x \in S$. Ahora bien, las condiciones de Kuhn-Tucker por hipótesis se satisfacen en x^* y por lo tanto, existen $u_i \geq 0$, $i=1,2,\dots,m$, no todos cero, tales que:

$$\text{grad}f(x^*) + \sum_{i \in I} u_i \text{grad}g_i(x^*) = 0$$

$$u_i g_i(x^*) = 0, \quad i \in I,$$

de donde

$$-\text{grad}f(x^*) = \sum_{i \in I} u_i \text{grad}g_i(x^*).$$

Pero $\sum_{i \in I} u_i \langle \text{grad}g_i(x^*), x-x^* \rangle \leq 0$ para toda $x \in S$, así

$$\sum_{i \in I} u_i \langle \text{grad}g_i(x^*), x-x^* \rangle = -\langle \text{grad}f(x^*), x-x^* \rangle \leq 0 \quad \text{para toda } x \in S, \text{ es}$$

decir $\langle \text{grad}f(x^*), x-x^* \rangle \geq 0$ para toda $x \in S$ y como f es pseudoconvexa, lo anterior implica que $f(x^*) \leq f(x)$ para toda $x \in S$, con lo cual se tiene el resultado pedido.

Obviamente, si las funciones f y g_i , $i=1,2,\dots,m$, son convexas, las condiciones de Kuhn-Tucker son también suficientes.

Considere ahora el problema:

(III). Minimizar $f(x)$

$$\text{suje}to \text{ a } g_i(x) \leq 0, \quad i=1,2,\dots,m$$

$$h_i(x) = 0, \quad i=m+1,\dots,p$$

$$x \in X,$$

donde X es un subconjunto abierto y no vacío de \mathbb{R}^n y las funciones f, g_i y h_i son de \mathbb{R}^n en \mathbb{R} . En este caso,

$$S = \{x \in X; g_i(x) \leq 0 \text{ para } i=1,2,\dots,m, \text{ y } h_i(x) = 0 \text{ para } i=m+1,\dots,p\}$$

Como una extensión natural del teorema 1.2.2.4 se tiene la siguiente condición geométrica necesaria de optimalidad para el problema (III).

Teorema. 1.2.2.8 Considere el problema (III). Sea $x^* \in X$ una solución factible y sea $I = \{i; g_i(x^*) = 0\}$. Suponga que para $i \notin I$, entonces g_i es continua en x^* y que las funciones f y g_i para $i \in I$ son diferenciables en el punto x^* . Si las funciones h_i , $i = m+1, \dots, p$, tienen derivada continua en x^* y si los vectores $\text{grad}h_i(x^*)$ son linealmente independientes en \mathbb{R}^n , $m+1 \leq i \leq p$ y

$$F(x^*) = \{d \neq 0; \langle \text{grad}f(x^*), d \rangle < 0\}$$

$$G(x^*) = \{d \neq 0; \langle \text{grad}g_i(x^*), d \rangle < 0 \text{ para } i \in I\}$$

$$H(x^*) = \{d \in \mathbb{R}^n; \langle \text{grad}h_i(x^*), d \rangle = 0 \text{ para } i = m+1, \dots, p\}$$

Entonces si x^* es una solución óptima local del problema (III).

$$F(x^*) \cap G(x^*) \cap H(x^*) = \emptyset.$$

Prueba:

Por contradicción, suponga que existe un vector $y \in F(x^*) \cap G(x^*) \cap H(x^*)$; esto es $\langle \text{grad}f(x^*), y \rangle < 0$, $\langle \text{grad}g_i(x^*), y \rangle < 0$ para toda $i \in I$ y $\langle \text{grad}h_i(x^*), y \rangle = 0$ para toda $i = m+1, \dots, p$. Ahora bien, para $\lambda \geq 0$ defina la función $\alpha: [0, \infty) \rightarrow \mathbb{R}^n$ mediante la ecuación $\frac{d}{d\lambda} \alpha(\lambda) = P(\lambda)y$ con la condición de frontera $\alpha(0) = x^*$, donde $P(\lambda)$ es la matriz que proyecta cualquier vector de \mathbb{R}^n en el espacio nulo de $\text{grad}H(\alpha(\lambda))$, donde $\text{grad}H(\alpha(\lambda))$ es la matriz con vectores columna $\text{grad}h_i(\alpha(\lambda))$, $i = m+1, \dots, p$. Para $\lambda > 0$ suficientemente pequeña, la ecuación diferencial anterior está bien definida y tiene solución, ya que, por hipótesis $\text{grad}H(x^*)$ tiene rango completo puesto que los vectores $\text{grad}h_i(x^*)$ son linealmente independientes en \mathbb{R}^n , también por hipótesis H tiene derivada continua en x^* , donde H es el vector que consiste de las funciones h_i , $i = m+1, \dots, p$, además la matriz P es continua en λ y $\alpha(\lambda) \rightarrow x^*$ cuando $\lambda \rightarrow 0^+$. A continuación se mostrará que para $\lambda > 0$ suficientemente pequeña $\alpha(\lambda)$ es un punto factible y $f(\alpha(\lambda)) < f(x^*)$, contradiciendo el hecho de que x^* es una solución óptima local del problema (III). Ahora bien, utilizando la regla de la cadena, se tiene que $\frac{d}{d\lambda} g_i(\alpha(\lambda)) = \text{grad}g_i(\alpha(\lambda)) \cdot \frac{d}{d\lambda} \alpha(\lambda) = \langle \text{grad}g_i(\alpha(\lambda)), P(\lambda)y \rangle$ para cada $i \in I$. En particular, el vector y está en el espacio nulo de $\text{grad}H(x^*)$, ya que,

por hipótesis $\langle \text{grad} h_i(x^*), y \rangle = 0$ para $i=m+1, \dots, p$, así para $\lambda=0$, $P(0)y=y$, $\frac{d}{d\lambda} g_i(\alpha(0)) = \langle \text{grad} g_i(x^*), y \rangle < 0$ para toda $i \in I$, de donde $g_i(\alpha(\lambda)) < 0$ para $\lambda > 0$ suficientemente pequeña, ya que, $\frac{d}{d\lambda} g_i(\alpha(0)) = \frac{g_i(\alpha(\lambda)) - g_i(\alpha(0))}{\lambda} + o(\lambda) = \frac{g_i(\alpha(\lambda)) - g_i(x^*)}{\lambda} + o(\lambda) = \frac{g_i(\alpha(\lambda))}{\lambda} + o(\lambda) < 0$, donde $\frac{|o(\lambda)|}{\lambda} \rightarrow 0$ cuando $\lambda \rightarrow 0$. Por otra parte, para $i \in I$, se tiene que $g_i(x^*) < 0$, y g_i es continua en x^* , de donde $g_i(\alpha(\lambda)) < 0$ para $\lambda > 0$ suficientemente pequeña. También puesto que el conjunto X es abierto, $\alpha(\lambda) \in X$ para $\lambda > 0$ suficientemente pequeña. Sólo falta mostrar que $h_i(\alpha(\lambda)) = 0$ para $\lambda > 0$ suficientemente pequeña para concluir que $\alpha(\lambda)$ para $\lambda > 0$ suficientemente pequeña es una solución factible del problema (III). Considerando el teorema del valor medio (1.1.3.12), se tiene que para algún $\mu \in (0, \lambda)$, se cumple que $h_i(\alpha(\lambda)) = h_i(\alpha(0)) + \lambda \frac{d}{d\lambda} h_i(\alpha(\mu)) = h_i(x^*) + \lambda \frac{d}{d\lambda} h_i(\alpha(\mu)) = \lambda \frac{d}{d\lambda} h_i(\alpha(\mu))$ para $i=m+1, \dots, p$. De nuevo por la regla de la cadena, se tiene que $\frac{d}{d\lambda} h_i(\alpha(\mu)) = \langle \text{grad} h_i(\alpha(\mu)), \frac{d}{d\lambda} \alpha(\mu) \rangle = \langle \text{grad} h_i(\alpha(\mu)), P(\mu)y \rangle$ para $i=m+1, \dots, p$. Pero por construcción $P(\mu)y$ está en el espacio nulo de $\text{grad} H(\alpha(\mu))$ y por lo tanto $\frac{d}{d\lambda} h_i(\alpha(\mu)) = 0$ para $i=m+1, \dots, p$ y como $h_i(\alpha(\lambda)) = \lambda \frac{d}{d\lambda} h_i(\alpha(\mu))$, se sigue que $h_i(\alpha(\lambda)) = 0$ para $i=1, 2, \dots, p$. Finalmente observe que $\frac{d}{d\lambda} f(\alpha(0)) = \langle \text{grad} f(x^*), y \rangle < 0$, de donde $f(\alpha(\lambda)) - f(x^*) < 0$ para $\lambda > 0$ suficientemente pequeña, como ya se había hecho notar, lo anterior contradice que x^* sea una solución óptima local del problema (III).

A continuación, se expresará la condición geométrica necesaria de optimalidad $F(x^*) \cap G(x^*) \cap H(x^*) = \emptyset$, en una forma algebraica más operacional, obteniéndose así las generalizaciones correspondientes a los teoremas 1.2.2.5, 1.2.2.6 y 1.2.2.7.

Teorema. 1.2.2.9 (Condiciones necesarias de Fritz John para el problema (III)).

Considere el problema (III). Sea $x^* \in X$ una solución factible y sea $I = \{i; g_i(x^*) = 0\}$. Suponga que las funciones g_i para $i \in \{1, 2, \dots, m\}$, así como la función f satisfacen ser diferenciables

en el punto $x^* \in X$. Suponga también que las funciones h_i para $i=m+1, \dots, p$ poseen derivada continua en el punto x^* . Si x^* es una solución óptima local del problema (III). Entonces existen u_0, u_i para $i \in I$ y v_i para $i=m+1, \dots, p$, tales que:

$$u_0 \text{grad}f(x^*) + \sum_{i=1}^m u_i \text{grad}g_i(x^*) + \sum_{i=m+1}^p v_i \text{grad}h_i(x^*) = 0,$$

$$u_0, u_i \geq 0, \quad i=1, 2, \dots, m$$

$$u_i g_i(x^*) = 0, \quad i=1, 2, \dots, m$$

Prueba:

Si los vectores $\text{grad}h_i(x^*)$ para $i=m+1, \dots, p$ son linealmente dependientes en \mathbb{R}^n , entonces existen escalares v_i , $i=m+1, \dots, p$, no todos cero, tales que $\sum_{i=m+1}^p v_i \text{grad}h_i(x^*) = 0$, entonces haciendo $u_0 = u_i = 0$ para $i=1, 2, \dots, m$, se tienen las condiciones pedidas. Suponga ahora que los vectores $\text{grad}h_i(x^*)$ para $i=m+1, \dots, p$ son linealmente independientes en \mathbb{R}^n . Sea T la matriz con renglones $\text{grad}f(x^*)^t$ y $\text{grad}g_i(x^*)^t$ para $i \in I$ y sea S la matriz con renglones $\text{grad}h_i(x^*)^t$ para $i=m+1, \dots, p$. Entonces por el teorema anterior el sistema $Td < 0, Sd = 0$ no tiene solución. Ahora considere los siguientes conjuntos $Q_1 = \{(z_1, z_2); z_1 = Td, z_2 = Sd\}$ y $Q_2 = \{(z_1, z_2); z_1 < 0 \text{ y } z_2 = 0\}$. Observe que Q_1 y Q_2 son conjuntos convexos, ajenos y no vacíos. Por el teorema 1.1.2.32, existe un vector $p = (u, v)$ no cero, tal que $u^t Td + v^t Sd \geq u^t z_1 + v^t z_2$ para todo $d \in \mathbb{R}^n$ y para todo $(z_1, z_2) \in Q_2$. Tomando $z_2 = 0$ y teniendo en cuenta que $z_1 < 0$ puede hacerse en valor absoluto arbitrariamente grande, se sigue que $u \geq 0$. También para $(z_1, z_2) = (0, 0) \in Q_2$, se tiene que $(u^t T + v^t S)d \geq 0$ y tomando $d = -(u^t T + v^t S)^t$, se sigue que $- \|T^t u + S^t v\|^2 \geq 0$, de donde $T^t u + S^t v = 0$. Por lo tanto se han obtenido u_0 y u_i para $i \in I$ no negativos, si $u = (u_0, u_1)$. Si se toman $u_i = 0$ para $i \notin I$, se concluye la prueba del teorema.

A continuación se presentan en dos teoremas las extensiones naturales de los resultados obtenidos en 1.2.2.6 y 1.2.2.7, sus pruebas se omiten, ya que, esencialmente no difieren de

de las pruebas de los teoremas 1.2.2.6 y 1.2.2.7, respectivamente.

Teorema. 1.2.2.10 (Condiciones necesarias de Kuhn-Tucker para el problema (III)).

Considere el problema (III). Sea $x^* \in X$ una solución factible y sea $I = \{i; g_i(x^*) = 0\}$. Suponga que las funciones f y g_i para $i \in \{1, 2, \dots, m\}$ son diferenciables en el punto $x^* \in X$ y que las h_i para $i \in \{m+1, \dots, p\}$ tienen derivada continua en el punto $x^* \in X$. Adicionalmente, suponga que los vectores $\text{grad}g_i(x^*)$ para $i \in I$ y $\text{grad}h_i(x^*)$ para $i = m+1, \dots, p$ son linealmente independientes en \mathbb{R}^n . Si x^* es una solución óptima local del problema (III). Entonces existen escalares u_i para $i \in \{1, 2, \dots, m\}$ y v_i para $i \in \{m+1, \dots, p\}$, tales que:

$$\text{grad}f(x^*) + \sum_{i=1}^m u_i \text{grad}g_i(x^*) + \sum_{i=m+1}^p v_i \text{grad}h_i(x^*) = 0$$

$$u_i g_i(x^*) = 0, \quad i = 1, 2, \dots, m$$

$$u_i \geq 0, \quad i = 1, 2, \dots, m.$$

Teorema. 1.2.2.11 (Condiciones suficientes de Kuhn-Tucker para el problema (III)).

Considere el problema (III). Sea $x^* \in X$ una solución factible y sea $I = \{i; g_i(x^*) = 0\}$. Suponga que la función f es pseudo-concava y que las funciones g_i para $i \in I$ son cuasi-concavas. Suponga que las condiciones de Kuhn-Tucker se cumplen en x^* , es to es, existen escalares $u_i \geq 0$ para $i \in I$ y v_i para $i \in \{m+1, \dots, p\}$, tales que:

$$\text{grad}f(x^*) + \sum_{i \in I} u_i \text{grad}g_i(x^*) + \sum_{j=m+1}^p v_j \text{grad}h_j(x^*) = 0.$$

Sea $T = \{i; v_i > 0\}$ y $S = \{i; v_i < 0\}$ y suponga además que las funciones h_i para $i \in T$ son cuasiconcavas y que las h_i para $i \in S$ son cuasi-concavas. Entonces x^* es una solución óptima global.

En lo que sigue se estudia el concepto de dualidad, el cual surge con regular frecuencia en optimización y en matemáticas en general. En abstracto, a cada ente matemático corresponde otro ente matemático que tiene ciertas propiedades en común con el original. En programación matemática este concepto aparece en forma natural.

Considere el siguiente problema de programación no lineal:

Minimizar $f(x)$
 sujeto a $g_i(x) \leq 0, i=1, 2, \dots, m$
 $h_i(x) = 0, i=m+1, \dots, p$
 $x \in X,$

donde $f, g_i, h_i: \mathbb{R}^n \rightarrow \mathbb{R}$ y X es un subconjunto no vacío de \mathbb{R}^n .

El problema dual lagrangeano asociado al problema planteado es definido como:

Maximizar $\theta(u, v)$
 sujeto a $v \geq 0,$
 donde $\theta(u, v) = \inf \left\{ f(x) + \sum_{i=1}^m u_i g_i(x) + \sum_{i=m+1}^p v_i h_i(x); x \in X \right\}.$

Note que la función lagrangeana dual $\theta(u, v)$ puede tomar el valor $-\infty$ para algún (u, v) y que en el dual las restricciones $g_i(x) \leq 0$ para $i=1, 2, \dots, m$ y $h_i(x) = 0$ para $i=m+1, \dots, p$ han sido incorporadas a la función objetivo mediante multiplicadores de Lagrange. Además, los multiplicadores de Lagrange asociados a las restricciones que son desigualdades están restringidos a no negatividad, mientras que los que están asociados a restricciones que son igualdades no tienen restricción en el signo. Finalmente, note que el problema dual lagrangeano es en realidad un problema max-min, ya que, se trata de maximizar el ínfimo de una función. Por esta última razón también se le conoce a veces como el problema dual max-min.

CAPITULO

2

FUNDAMENTACION DE LA TEORIA
DE ALGORITMOS DE
PROGRAMACION NO LINEAL

CAPITULO 2. FUNDAMENTACION DE LA TEORIA DE ALGORITMOS DE PROGRAMACION NO LINEAL

2.1 INTRODUCCION

Un hecho significativo para el desarrollo de una teoría general de algoritmos de programación no lineal, consiste en asociar a cada algoritmo de solución un mapeo de punto a conjunto; Un algoritmo de solución será entendido como un procedimiento iterativo, el cual genera una sucesión de puntos de acuerdo a un conjunto de instrucciones preestablecidas y a la información cedida por puntos previamente obtenidos. Los conceptos que más han destacado explotando la asociación de cada algoritmo de solución con un mapeo de punto a conjunto para la obtención de resultados generales sobre convergencia algorítmica, son los conceptos de función de descenso y de mapeo cerrado, este último está íntimamente relacionado con cierta propiedad de continuidad que poseen los algoritmos convergentes.

2.2 MAPEOS ALGORITMICOS Y FUNCIONES DE DESCENSO

En esta sección se proporcionan los fundamentos necesarios para el desarrollo de una teoría de algoritmos de programación no lineal. En primer término, se presentan los conceptos de mapeo de punto a conjunto y de mapeo algorítmico. Después se describen los conjuntos solución más usuales. Por último, se introducen las nociones de convergencia algorítmica y de función de descenso.

Antes de introducir la noción de mapeo algorítmico, es necesario generalizar el concepto de función en el siguiente sentido:

Definición. 2.2.1 Sean X y Y dos conjuntos no vacíos. Se dice que un mapeo $T: X \rightarrow 2^Y$ es de punto a conjunto, si dado $x \in X$, se tiene que $T(x) \in 2^Y$, equivalentemente que $T(x) \subset Y$. (2^Y es el conjunto potencia de Y).

El hecho de que $T: X \rightarrow 2^Y$, se denotará simplemente por $T: X \rightarrow Y$ es un mapeo de punto a conjunto.

Ejemplo. 2.2.2 Dado $x \in \mathbb{R}$, sea $T(x) = \{y; -|1-x| < y < |2-x|\}$, entonces $T: \mathbb{R} \rightarrow \mathbb{R}$ es un mapeo de punto a conjunto.

Un algoritmo será considerado como un mapeo de punto a conjunto aplicado en forma recursiva de tal manera que se genere una sucesión de puntos. Más precisamente, dado un vector $x_k \in \mathbb{R}^n$, si se aplican las instrucciones del algoritmo se obtiene un nuevo punto x_{k+1} . Este proceso puede entonces ser descrito por un mapeo de punto a conjunto T , mediante la relación $x_{k+1} \in T(x_k)$.

Definición. 2.2.3 Un algoritmo es un mapeo de punto a conjunto $T: X \rightarrow X$, $X \subset \mathbb{R}^n$ que aplicado en forma recursiva genera una sucesión infinita de puntos.

Los términos algoritmo y mapeo algorítmico serán utilizados indistintamente.

Observe que cada algoritmo tiene asociado un mapeo de punto a conjunto. Sin embargo el recíproco de esta proposición no es cierto. En efecto, sea $T: \mathbb{R} \rightarrow \mathbb{R}$ un mapeo de punto a conjunto definido por $T(x) = \{y; y < x \text{ y } y > x\}$ para cada $x \in \mathbb{R}$, entonces T satisface $T(x) = \emptyset$ para todo $x \in \mathbb{R}$. Teniéndose como consecuencia que T no genera sucesiones de puntos. Por lo tanto T no define un algoritmo.

Otro caso extremo que pudiera no tener sentido con lo que intuitivamente se entiende por un algoritmo, es que si se

define el mapeo de punto a conjunto $T: \mathbb{R} \rightarrow \mathbb{R}$ como $T(x) = \{c\}$ para todo $x \in \mathbb{R}$, entonces se tendrá que $x_k = c$ para toda $k \in \mathbb{N}$ y sin embargo, por definición T es un algoritmo.

Ejemplo. 2.2.4 Sea $T: \mathbb{R} \rightarrow \mathbb{R}$ el mapeo de punto a conjunto de finido en el ejemplo 2.2.2. Sea $x_1 = 2$, entonces $T(x_1) = T(2) = \{y; -1 < y < 0\}$, si x_2 ha de pertenecer a $T(x_1)$, ponga por ejemplo que $x_2 = -\frac{1}{2}$. Así $T(x_2) = T(-\frac{1}{2}) = \{y; -\frac{1}{2} < y < 3/2\}$, si x_3 ha de estar en $T(x_2)$, ponga por ejemplo que $x_3 = 1$, etc. Entonces $\{x\}_{k \in \mathbb{N}} = \{2, -\frac{1}{2}, 1, \dots\}$.

El propósito que lleva consigo el ejemplo anterior, es el de mostrar que la expresión $x_{k+1} \in T(x_k)$ no debe ser confundida con $x_{k+1} = T(x_k)$.

A primera vista, podría parecer que la definición de mapeo algorítmico es superflua, ya que en la práctica no se escogen puntos arbitrarios pertenecientes a un conjunto determinado, sino que se tiene identificado un punto en cada iteración del procedimiento de solución. Sin embargo, si un algoritmo se programa y se corre dos veces en una computadora digital con diferentes puntos iniciales, seguramente se generarán dos sucesiones diferentes. Es aquí en donde se destaca la importancia del concepto de mapeo algorítmico, pudiéndose relacionar las dos sucesiones como generadas por el mismo mapeo algorítmico.

A continuación se presenta un ejemplo, en el cual se muestran las ideas establecidas:

Ejemplo. 2.2.5 Considere el problema de programación no lineal

$$\text{Minimizar } f(x) = x^2 \quad (f: \mathbb{R} \rightarrow \mathbb{R})$$

sujeto a $x \geq 1$.

Obviamente la solución óptima global se tiene para $x^* = 1$. Sea $T: \mathbb{R} \rightarrow \mathbb{R}$ el mapeo algorítmico definido por $T(x) = \{\frac{1}{2}(x+1)\}$.

Puede ser fácilmente verificado que la sucesión generada por la aplicación repetida de T con cualquier punto inicial converge a la solución óptima $x^*=1$. En efecto, note que la aplicación repetida de T conduce a la relación

$$x_{k+1} \in T(x_k) = T^k(x_1) = \left\{ \frac{x_1}{2^k} + ((1/2) + (1/4) + (1/8) + \dots + (1/2^k)) \right\}$$

y claramente $x_{k+1} \rightarrow 1$ cuando $k \rightarrow \infty$.

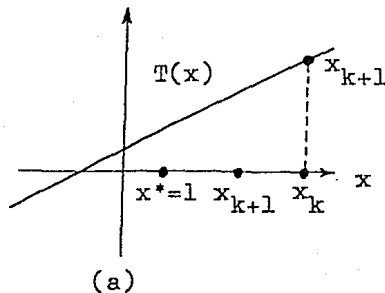
Ahora considere el mapeo algorítmico $S: \mathbb{R} \rightarrow \mathbb{R}$, definido por

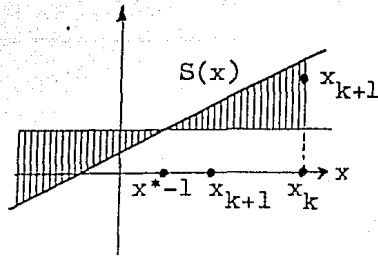
$$S(x) = \begin{cases} [1, (1/2)(x+1)] & \text{si } x \geq 1 \\ [(1/2)(x+1), 1] & \text{si } x < 1 \end{cases}$$

La imagen $S(x)$ de cualquier $x \in \mathbb{R}$ es un intervalo cerrado y cualquier punto en ese intervalo puede ser seleccionado en calidad de sucesor de x . Al igual que antes, independientemente del punto inicial, el mapeo algorítmico S converge también a la solución óptima $x^*=1$. Para $x_1=4$, las sucesiones generadas por T y S , respectivamente son:

$$\{4, 2.5, 1.75, 1.375, 1.1875, \dots\} \text{ y } \{4, 2, 1.2, 1.1, 1.02, \dots\}$$

Note también que tanto T como S tienen como dominio a \mathbb{R} y no al conjunto definido por la restricción $x \geq 1$. Teniéndose con ésto que un mapeo algorítmico no involucra condiciones de factibilidad necesariamente.





(b)

Figura 28. Los incisos (a) y (b) son los mapeos algorítmicos del ejemplo 2.2.5

A continuación se describirán los conjuntos solución más usuales en programación no lineal.

Considere el problema de programación no lineal

Minimizar $f(x)$

sujeto a $x \in S$,

donde $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $S \subset \mathbb{R}^n$, $S \neq \emptyset$.

Es claro pues, que la aplicación de un algoritmo al problema planteado, tiene como propósito encontrar un punto que cumpla con ciertos requisitos. En nuestro caso estos requisitos son condiciones de optimalidad.

Una propiedad deseable que un algoritmo debe poseer, es que éste genere una sucesión de puntos que converja a una solución óptima del problema. Sin embargo, en muchos casos basta con obtener puntos suficientemente próximos a dicha solución óptima, por ejemplo, cuando se carece de propiedades de convexidad, o bien por el tamaño del problema, etc., en cuyo caso el proce-

dimiento iterativo debe ser detenido si el punto obtenido pertenece a un conjunto predeterminado Ω , el cual será llamado conjunto solución. En seguida, se listan algunos de los conjuntos solución más usuales en programación no lineal:

- (i). $\Omega = \{x \in \mathbb{R}^n; x \text{ es una solución óptima local}\}$
- (ii). $\Omega = \{x \in \mathbb{R}^n; x \in S \text{ y } f(x) \leq b\}$, donde b es un valor aceptable de la función objetivo.
- (iii). $\Omega = \{x \in \mathbb{R}^n; x \in S \text{ y } f(x) < L + \epsilon\}$, donde $\epsilon > 0$ es una tolerancia especificada y L es una cota inferior de la solución óptima.
- (iv). $\Omega = \{x \in \mathbb{R}^n; x \in S \text{ y } f(x) - f(x^*) < \epsilon\}$, donde $f(x^*)$ es la solución óptima global y $\epsilon > 0$ una tolerancia especificada.
- (v). $\Omega = \{x \in \mathbb{R}^n; x \in S \text{ y satisface las condiciones de optimalidad de Kuhn-Tucker}\}$.
- (vi). $\Omega = \{x \in \mathbb{R}^n; x \in S \text{ y satisface las condiciones de Fritz Jhon}\}$.

En general, la convergencia de algoritmos es basada en la noción de conjunto solución.

Definición. 2.2.6 Se dice que un mapeo algorítmico $T: X \rightarrow X$, $X \subset \mathbb{R}^n$ converge sobre $Y \subset X$ a un punto en Ω , si al comenzar con cualquier punto inicial $x_1 \in Y$, el límite de cualquier subsucesión convergente $\{x_n\}_{m \in \mathbb{N}}$ de la sucesión $\{x_n\}_{n \in \mathbb{N}}$ generada por el mapeo algorítmico, pertenece al conjunto solución Ω .

Ejemplo. 2.2.7 Considere los dos mapeos algorítmicos T y S del ejemplo 2.2.5, entonces los mapeos algorítmicos T y S convergen sobre \mathbb{R} a un punto en Ω , ya que, la aplicación repetida de T ó S con cualquier punto inicial converge a la solución óptima $x^* = 1$ y por lo tanto cualquier subsucesión de las sucesiones generadas por T ó S converge también a $x^* = 1$.

Una vez que han sido introducidos los conceptos de mapeo algorítmico, conjunto solución y convergencia algorítmica, se precisará la noción de función de descenso en un algoritmo.

Definición. 2.2.8 Sea $T: X \rightarrow X$, $X \subset \mathbb{R}^n$ un mapeo algorítmico y sea $\Omega \subset X$ un conjunto solución. Se dice que una función continua $F: X \rightarrow \mathbb{R}$ es una función de descenso para T y Ω , si F satisface las siguientes condiciones:

- (i). $x \notin \Omega$ y $y \in Tx$ implica $F(y) < F(x)$
- (ii). $x \in \Omega$ y $y \in Tx$ implica $F(y) \leq F(x)$.

Casi siempre, la función objetivo misma, sirve como función de descenso. Sin embargo, esta regla no es general. Por ejemplo, en problemas irrestrictos de programación no lineal, una función de descenso podría ser $F(x) = \|\text{grad}f(x)\|$ si se define $\Omega = \{x \in \mathbb{R}^n; \text{grad}f(x) = 0\}$. Esto se muestra en el siguiente ejemplo:

Ejemplo. 2.2.9 Considere el problema irrestricto de programación no lineal

Minimizar $f(x) = x^2$ ($f: [0, \infty) \rightarrow [0, \infty)$)

y considere el mapeo algorítmico $T: [0, \infty) \rightarrow [0, \infty)$, definido por $T(x) = [0, \frac{1}{2}x]$, sea $\Omega = \{0\}$ y sea $F: [0, \infty) \rightarrow \mathbb{R}$, $F(x) = \|\text{grad}f(x)\|$. Para toda $y \in T(x)$, se tiene que cuando $x \neq 0$, se satisface $0 < y < x$, así

$$F(y) = \|\text{grad}f(y)\| = 2y < 2x = \|\text{grad}f(x)\| = F(x).$$

Ahora bien, si $x \in \Omega$, es decir, $x = 0$, y si $y \in T(0) = \{0\}$, entonces $F(y) = 0 = F(x)$. En consecuencia $F(\cdot) = \|\text{grad}f(\cdot)\|$ es una función de descenso para T y Ω .

Definición. 2.2.10 Sea $T: X \rightarrow X$, $X \subset \mathbb{R}^n$ un mapeo algorítmico y sea $\Omega \subset X$ un conjunto solución. Se dice que el algoritmo T es de descenso, si existe una función de descenso para T y Ω .

2.3 MAPEOS CERRADOS

En esta sección se introducirá la noción de mapeo cerrado de punto a conjunto. El significado de la cerradura será estudiado mediante algunos ejemplos y una discusión subsecuente. El tener una medida de mejoramiento en un algoritmo, como puede ser la función de descenso, es una propiedad deseable. No obstante, lo anterior no es suficiente para garantizar que el algoritmo converja a un punto perteneciente al conjunto solución Ω . Ahora bien, todos los algoritmos convergentes poseen cierta propiedad de continuidad, que es quizás la más significativa para la obtención de resultados generales sobre convergencia algorítmica. Dicha propiedad de continuidad, como se verá está relacionada con la noción de cerradura de un mapeo algorítmico.

A continuación se presenta un conjunto de ejemplos, que tienen la finalidad de mostrar la necesidad de introducir la noción de cerradura en un mapeo algorítmico como una propiedad de continuidad para alcanzar convergencia algorítmica.

Ejemplo. 2.3.1 Considere el problema

Maximizar $f(x)=5-x$ ($f:\mathbb{R} \rightarrow \mathbb{R}$)

sujeto a $x \geq 0$.

Obviamente la solución óptima global del problema planteado es $x^*=0$. Sea $\Omega=\{0\}$ y sea $T:\mathbb{R} \rightarrow \mathbb{R}$ el mapeo algorítmico definido por:

$$Tx = \begin{cases} \{1+\frac{1}{2}(x-1)\}, & \text{si } x > 1 \\ \{\frac{1}{2}x\} & , \text{ si } x \leq 1 \end{cases}$$

Defina la función de ascenso como: $F(x)=f(x)=5-x$. Entonces para $x_1=5$, el algoritmo genera la sucesión:

$$\{x_k\}_{k \in \mathbb{N}} = \{5, 3, 2, 3/2, 5/4, 9/8, 17/16, \dots\}$$

Así,

$$\{F(x_k)\}_{k \in \mathbb{N}} = \{0, 2, 3, 7/2, 15/4, \dots\}$$

Note que $F(x_{k+1}) > F(x_k)$ para toda $k \in \mathbb{N}$, lo cual muestra que efectivamente F es una función de ascenso para T y Ω . Sin embargo, $x_k \rightarrow 1 \neq x^* = 0$ cuando $k \rightarrow \infty$, ya que, la relación $x_{k+1} \in T(x_k) = T^k(x_1) = \{1 + (\frac{1}{2})^k(x_1 - 1)\}$ implica que $x_{k+1} \rightarrow 1$ cuando $k \rightarrow \infty$ y desde luego $1 \notin \Omega$.

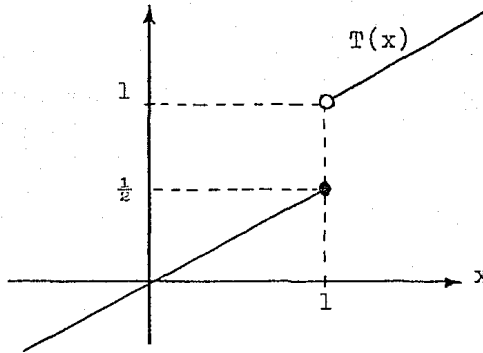


Figura 29. Mapeo algorítmico del ejemplo 2.3.1

Si se examina el mapeo algorítmico definido en el ejemplo anterior. Se observará que no es continuo en el punto $x=1$. Esta discontinuidad es precisamente la que impide la convergencia a la solución óptima. Algo más, aunque F es una función de ascenso para T y Ω , se tiene que la existencia de dicha F no garantiza la convergencia de T a un punto de Ω .

Ejemplo. 2.3.2 Considere el problema

$$\begin{aligned} &\text{Minimizar } f(x) = x^2 \quad (f: \mathbb{R} \rightarrow \mathbb{R}) \\ &\text{sujeto a } x \geq 1. \end{aligned}$$

Claramente la solución óptima global es el punto $x^*=1$. Defina $\Omega = \{1\}$ y considere el mapeo algorítmico $T: \mathbb{R} \rightarrow \mathbb{R}$ definido por:

$$Tx = \begin{cases} [(3/2) + \frac{1}{2}x, 1 + \frac{1}{2}x], & \text{si } x \geq 2 \\ \{\frac{1}{2}(x+1)\} & , \text{ si } x < 2 \end{cases}$$

Para cualquier punto inicial $x_1 \geq 2$, se tiene que la sucesión generada por T converge a $\bar{x}=2$. Note que $\bar{x} \notin \Omega$. Por otra parte, para cualquier $x_1 < 2$, la sucesión generada por T converge a $x^*=1$. En este ejemplo el algoritmo T converge sobre el intervalo $(-\infty, 2)$ a un punto de Ω y no converge sobre el intervalo $[2, \infty)$ a un punto en Ω .

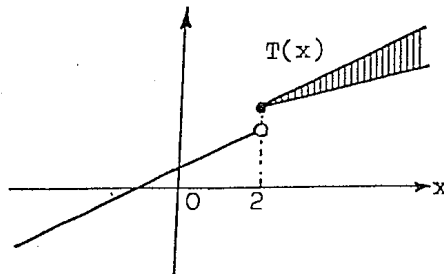


Figura 30. Ejemplo de un mapeo algorítmico no convergente.

Observe que los mapeos algorítmicos de los ejemplos 2.2.5 y 2.3.2 poseen las siguientes propiedades:

- (i). Dado un punto factible x_k , se tiene que el punto x_{k+1} es también factible.
- (ii). Dado un punto factible $x_k \notin \Omega$, el punto sucesor satisface $f(x_{k+1}) < f(x_k)$, en otras palabras, la función f es estrictamen

te decreciente.

(iii). Dado un punto factible $x_k \in \Omega$, se cumple que los puntos sucesores pertenecen también a Ω .

A pesar de la similitud, los dos algoritmos del ejemplo 2.2.5 convergen a $x^*=1$. Mientras que el algoritmo del ejemplo 2.2.3 no converge a $x^*=1$ si el punto inicial x_1 , se elige de tal manera que $x_1 \geq 2$.

La noción que se explotará para estudiar la continuidad de mapeos algorítmicos, es la noción de mapeo cerrado de punto a conjunto que a continuación se presenta:

Definición. 2.3.3 Sean X y Y dos subconjuntos cerrados y no vacíos de \mathbb{R}^n . Sea $T: X \rightarrow Y$ un mapeo de punto a conjunto. Se dice que el mapeo T es cerrado en un punto $x \in X$ si cuando:

(i). $\{x_k\}_{k \in \mathbb{N}}$ es una sucesión de elementos de X , tal que $x_k \rightarrow x^*$ cuando $k \rightarrow \infty$.

(ii) y si $\{y_k\}_{k \in \mathbb{N}}$ es una sucesión de elementos de Y , tal que $y_k \in Tx_k$ y $y_k \rightarrow y^*$ cuando $k \rightarrow \infty$.

Se implica que $y^* \in Tx^*$.

Un mapeo de punto a conjunto T es llamado cerrado sobre $Z \subset X$ si T es cerrado en todo punto de Z .

Ejemplo. 2.3.4 Considere el mapeo algorítmico T del ejemplo 2.3.2. T no es un mapeo cerrado en el punto $x^*=2$. En efecto, la sucesión $x_k = 2 - \frac{1}{k} \rightarrow x^*=2$, sin embargo, la sucesión $y_k = \frac{1}{2}(x_k + 1) = \frac{1}{2}(2 - \frac{1}{k} + 1) = \frac{3}{2} - \frac{1}{2k} \rightarrow y^* = \frac{3}{2}$ y $y^* \notin Tx^* = \{2\}$.

Ejemplo. 2.3.5 Defina $T: [0, \infty) \rightarrow [0, \infty)$, mediante $Tx = \{y; \frac{1}{3}x \leq y \leq x^2\}$. Se verá que T es un mapeo cerrado sobre $[0, \infty)$. Es decir, T es un mapeo cerrado en todo punto $x \geq 0$.

Sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión de números no negativos, tal que $x_k \rightarrow x^*$ cuando $k \rightarrow \infty$ y sea $\{y_k\}_{k \in \mathbb{N}}$ con $y_k \in Tx_k$ y $y_k \rightarrow y^*$

Ahora bien, $y_k \in Tx_k$ implica que $\frac{1}{3}x_k \leq y_k \leq \frac{2}{3}x_k$ para toda $k \in \mathbb{N}$, de donde $\frac{1}{3} \lim_{k \rightarrow \infty} x_k \leq \lim_{k \rightarrow \infty} y_k \leq \frac{2}{3} \lim_{k \rightarrow \infty} x_k$, así $\frac{1}{3}x^* \leq y^* \leq \frac{2}{3}x^*$, teniéndose que $y^* \in Tx^*$. Algo más, si se define $\Omega = \{0\}$ y se toma cualquier $x_1 \in (0, \omega)$, se tiene que $x_{k+1} \in Tx_k$ implica $0 < x_{k+1} < x_k$. Por lo tanto $\{x_k\}_{k \in \mathbb{N}}$ es una sucesión monótona decreciente y acotada, en consecuencia es convergente. Es fácil ver que converge a cero, ya que, $\frac{1}{3}x_k \leq y_k \leq \frac{2}{3}x_k$ implica $(\frac{1}{3})^k x_1 \leq x_{k+1} \leq (\frac{2}{3})^k x_1$, de donde $\lim_{k \rightarrow \infty} x_k = 0$, en cuyo caso se tiene que cualquier subsucesión converge también a cero y por lo tanto el mapeo algorítmico T converge sobre $[0, \omega)$ a un punto de Ω .

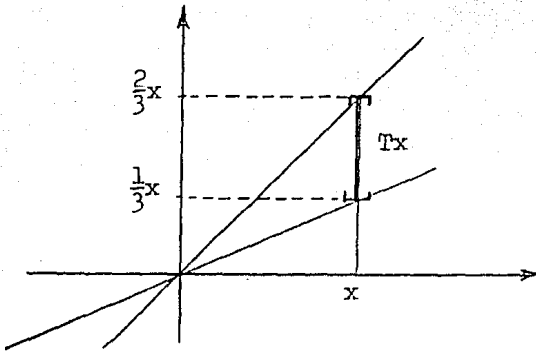


Figura 31. Mapeo algorítmico del ejemplo 2.3.5

Finalmente, se mostrará mediante un ejemplo, que el hecho de tener un mapeo cerrado no garantiza la convergencia algorítmica.

Ejemplo. 2.3.6 Considere el mapeo algorítmico $T: \mathbb{R} \rightarrow \mathbb{R}$ definido por $Tx = \{y; -1 \leq y \leq 1\}$ para toda $x \in \mathbb{R}$. Sea $\Omega = (-z, z)$ con $z \in (0, 1)$,

entonces $x_k = (-1)^k$ define una sucesión infinita que puede ser generada por la aplicación repetida de T , esta sucesión tiene dos subsucesiones infinitas convergentes $\{-1, -1, -1, \dots\}$ y $\{1, 1, 1, \dots\}$, sin embargo, los puntos $-1, 1 \notin \Omega$. Por lo tanto el mapeo algorítmico T no converge sobre \mathbb{R} a un punto de Ω .

CAPITULO

3

CONVERGENCIA DE MAPEOS
ALGORITMICOS

CAPITULO 3. CONVERGENCIA DE MAPEOS ALGORITMICOS

3.1 INTRODUCCION

En 1970 Zangwill y Polak presentaron por primera vez un conjunto de condiciones suficientes para alcanzar convergencia algorítmica, motivando con esto, el desarrollo de una teoría general de algoritmos de programación no lineal. Recientemente, Meyer y Komlósi han obtenido resultados más generales explotando la noción de compacidad.

En el transcurso de este capítulo se establecen las propiedades deseables que un algoritmo debe poseer. En general, los algoritmos generan sucesiones infinitas de puntos, siendo posible que ninguno de ellos califique estrictamente como una solución óptima. Se probará pues, que un algoritmo con las propiedades establecidas, efectivamente converge a un punto del conjunto solución.

También, en este capítulo se estudiará la convergencia de composiciones de mapeos algorítmicos, ya que, muchos de los algoritmos desarrollados en programación no lineal, por su complejidad, pueden ser considerados como composición de mapeos algorítmicos. Por último, se presentará una discusión sobre la rapidez de convergencia de un algoritmo.

3.2 EL TEOREMA FUNDAMENTAL DE CONVERGENCIA ALGORITMICA

Una vez que han sido introducidos los conceptos de mapeo algorítmico, convergencia algorítmica y mapeo cerrado, se presentará un conjunto de condiciones suficientes para la convergencia de un mapeo algorítmico. Para lo cual, es necesario

tener en cuenta el concepto de compacidad estudiado en el capítulo 1. Dichas condiciones serán establecidas mediante el teorema fundamental de convergencia algorítmica.

Se comenzará por probar el siguiente lema:

Lema. 3.2.1 Sea $X \subset \mathbb{R}^n$, $X \neq \emptyset$, sea $F: X \rightarrow \mathbb{R}$ una función continua, y sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión infinita de puntos en X , tales que:

(i). $F(x_k) < F(x_{k+1})$ para todo $k \in \mathbb{N}$.

(ii). Existe una subsucesión $\{x_{k_m}\}_{m \in \mathbb{N}}$ de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ tal que $x_{k_m} \rightarrow x^*$ cuando $m \rightarrow \infty$.

Entonces $\lim_{k \rightarrow \infty} F(x_k) = \lim_{m \rightarrow \infty} F(x_{k_m}) = F(x^*)$.

Prueba:

En primer lugar, como F es continua, entonces $\lim_{m \rightarrow \infty} F(x_{k_m}) = F(x^*)$. Así, dado $\epsilon > 0$, existe un $N \in \mathbb{N}$, tal que $F(x^*) - F(x_{k_m}) = |F(x^*) - F(x_{k_m})| < \epsilon$ para toda $m \in \mathbb{N}$ que satisface $m > N$. Note que la sucesión $\{F(x_{k_m})\}_{m \in \mathbb{N}}$ es monótona creciente y convergente, teniéndose así que converge al supremo. En otras palabras $F(x^*) = \sup_m \{F(x_{k_m})\} \geq F(x_{k_m})$ para toda $m \in \mathbb{N}$, de donde $F(x^*) - F(x_{k_m}) \geq 0$ para toda $m \in \mathbb{N}$. En particular, debido a la convergencia de la sucesión $\{F(x_{k_m})\}_{m \in \mathbb{N}}$, $F(x^*) - F(x_N) < \epsilon$.

También a consecuencia de la hipótesis (i), se tiene que $F(x_N) \leq F(x_k)$ para toda $k > N$. Así,

$F(x^*) - F(x_k) = (F(x^*) - F(x_N)) + (F(x_N) - F(x_k)) < \epsilon + 0 = \epsilon$ para toda $k > N$. Por lo tanto, $\lim_{k \rightarrow \infty} F(x_k) = F(x^*)$, con lo que se concluye la prueba del lema.

Teorema. 3.2.2 (Teorema fundamental de convergencia algorítmica)

Sea X un subconjunto cerrado y no vacío de \mathbb{R}^n , sea $\Omega \neq \emptyset$, $\Omega \subset X$ un conjunto solución y sea $T: X \rightarrow X$ un mapeo algorítmico. Dado $x_1 \in X$, la sucesión $\{x_k\}_{k \in \mathbb{N}}$ es generada como sigue: Si $x_k \in \Omega$ se detiene el algoritmo; En caso contrario, tome $x_{k+1} \in Tx_k$, reemplace k por $k+1$ e itere de nuevo.

Suponga que la sucesión generada por el mapeo algorítmico T , $\{x_k\}_{k \in \mathbb{N}}$ está totalmente contenida en un conjunto compacto K de X y suponga que existe una función F de ascenso para T y Ω . Si T es un mapeo cerrado sobre $X \setminus \Omega$. Entonces el mapeo algorítmico converge sobre X , es decir,

(i). El algoritmo se detiene en un número finito de pasos con un punto en Ω , ó

(ii). Cada subsucesión convergente $\{x_{k_m}\}_{m \in \mathbb{N}}$ de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ converge a punto en el conjunto solución Ω , esto es, todos los puntos de acumulación de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ pertenecen a Ω .

Prueba:

Si en alguna iteración se genera un punto en Ω , entonces el algoritmo se detiene. Suponga ahora, que una sucesión infinita es generada. Sea $\{x_{k_m}\}_{m \in \mathbb{N}}$ cualquier subsucesión convergente a un punto $x^* \in X$. Por el lema anterior, se tiene que $\lim_{k \rightarrow \infty} F(x_k) = \lim_{m \rightarrow \infty} F(x_{k_m}) = F(x^*)$. Para demostrar que $x^* \in \Omega$, suponga lo contrario, es decir, $x^* \notin \Omega$. Considere la sucesión $\{x_{k_m+1}\}_{m \in \mathbb{N}}$, ésta sucesión está contenida en el conjunto compacto $K \subset X$ y por lo tanto contiene una subsucesión convergente $\{x_{k_{m_p}+1}\}_{p \in \mathbb{N}}$ con límite $\bar{x} \in X$. De nuevo por el lema anterior

$$F(\bar{x}) = \lim_{p \rightarrow \infty} F(x_{k_{m_p}+1}) = \lim_{k \rightarrow \infty} F(x_k) = F(x^*)$$

Ahora bien, como $x_{k_m p} \rightarrow x^*$ cuando $p \rightarrow \infty$, ya que, $\{x_{k_m p}\}_{p \in \mathbb{N}}$ es una subsucesión de la sucesión $\{x_{k_m}\}_{m \in \mathbb{N}}$ la cual es convergente a x^* y como $x_{k_m p+1} \in Tx_{k_m p}$, se sigue que $\bar{x} \in Tx^*$ por ser T cerrado en $X \setminus \Omega$, pero observe que $x^* \notin \Omega$, de donde $F(x^*) < F(\bar{x})$, lo cual contradice el hecho de que $F(\bar{x}) = F(x^*)$. En consecuencia $x^* \in \Omega$.

El teorema anterior muestra que además de una medida de progreso definida por una función de ascenso y una propiedad de continuidad debida a la cerradura de T , se requiere además que la sucesión generada por el mapeo algorítmico T esté contenida en un conjunto compacto, con la finalidad de asegurar que todos los puntos de acumulación de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ pertenecen al conjunto solución Ω .

Ejemplo. 3.2.3 Sea $T: [0, \infty) \rightarrow [0, \infty)$ el mapeo de punto a conjunto del ejemplo 2.3.5. Si $\Omega = \{0\}$, se vió que T era un mapeo cerrado en $[0, \infty)$ y convergente sobre $[0, \infty)$ a un punto de Ω . Note que para cualquier $x_1 \in [0, \infty)$, se tiene que $x_k \in [0, x_1] = KC [0, \infty)$ para todo $k \in \mathbb{N}$, ya que $x_{k+1} \leq x_k$ para toda $k \in \mathbb{N}$. Así K es un conjunto compacto de $[0, \infty)$ para cualquier $x_1 \geq 0$.

Corolario. 3.2.4 Bajo las hipótesis del teorema anterior, si Ω consiste de un sólo punto, digamos, $\Omega = \{x^*\}$. Entonces la sucesión $\{x_k\}_{k \in \mathbb{N}}$ generada por el mapeo algorítmico T converge a x^* .

Prueba:

Suponga lo contrario, es decir, que existe un $\epsilon > 0$ y una subsucesión $\{x_{k_m}\}_{m \in \mathbb{N}}$ de la sucesión $\{x_k\}_{k \in \mathbb{N}}$, tal que $\|x_{k_m} - x^*\| > \epsilon$ para todo $m \in \mathbb{N}$. Observe que existe una subsucesión $\{x_{k_m p}\}_{p \in \mathbb{N}}$

de la sucesión $\{x_{k_m}\}_{m \in \mathbb{N}}$, tal que $x_{k_m} \rightarrow \bar{x} \in X$ debido a la hipótesis de compacidad. Por el teorema anterior se tiene que $\bar{x} \in \Omega$. Pero $\Omega = \{x^*\}$, teniéndose así que $\bar{x} = x^*$. Por lo tanto, $x_{k_m} \rightarrow x^*$, es decir, existe un $N \in \mathbb{N}$ tal que $\|x_{k_m} - x^*\| < \epsilon$ para toda $p > N$, pero si $\{x_{k_m_p}\}_{p \in \mathbb{N}}$ es una subsucesión de $\{x_{k_m}\}_{m \in \mathbb{N}}$, entonces también $\|x_{k_m_p} - x^*\| > \epsilon$ para todo $p \in \mathbb{N}$. Esta contradicción prueba el corolario.

De acuerdo al teorema 3.2.3, un algoritmo es detenido cuando se obtiene un punto $x \in \Omega$ a partir de la sucesión generada por él. En muchos casos, sin embargo, la convergencia a un punto en el conjunto solución Ω sólo ocurre en el límite, teniendo que recurrir a reglas prácticas para detener el procedimiento iterativo del algoritmo. A continuación, se presentan las reglas usadas con más frecuencia para detener un algoritmo:

Sean $\epsilon > 0$ y N un número natural predeterminados, los siguientes criterios son frecuentemente utilizados para detener un algoritmo:

(i). $\|x_{k+N} - x_k\| = \|T^N(x_k) - x_k\| < \epsilon$

Aquí, el algoritmo es detenido si el desplazamiento después de N aplicaciones consecutivas del mapeo algorítmico T es menor que ϵ .

(ii). $\frac{\|x_{k+N} - x_k\|}{\|x_k\|} = \frac{\|T^N(x_k) - x_k\|}{\|x_k\|}, \quad x_k \neq 0$

Bajo este criterio, el algoritmo es detenido si el desplazamiento relativo con respecto a la k -ésima iteración después de N aplicaciones consecutivas del mapeo algorítmico T es menor que ϵ .

(iii). $|F(x_{k+N}) - F(x_k)| < \epsilon$

Aquí, el algoritmo es detenido si el mejoramiento total en

la función de ascenso (descenso) después de N aplicaciones consecutivas del mapeo algorítmico T es menor que ϵ .

$$(iv). \frac{|F(x_{k+N}) - F(x_k)|}{|F(x_k)|} < \epsilon, F(x_k) \neq 0$$

Si el mejoramiento relativo de la función de ascenso (descenso) con respecto a la iteración k después de N aplicaciones consecutivas del mapeo algorítmico T es menor que ϵ , entonces el algoritmo es detenido.

3.3 COMPOSICION DE MAPEOS

Muchos de los algoritmos desarrollados en programación no lineal son bastante complejos, estos algoritmos frecuentemente son la composición de varios mapeos algorítmicos. Por ejemplo, algunos algoritmos determinan primero una dirección factible d_k para desplazarse a través de ella y posteriormente determinan la longitud del paso λ_k resolviendo el problema unidimensional de minimizar $F(x_k - \lambda d_k)$, donde F es una función de descenso. En este caso el procedimiento de solución es la composición de los mapeos algorítmicos T_1 y T_2 ($T_2 \circ T_1$), donde T_1 determina la dirección factible d_k y T_2 encuentra la solución óptima del problema unidimensional de minimizar $F(x_k - \lambda d_k)$. Para formalizar este tipo de situaciones se proporciona la siguiente definición:

Definición. 3.3.1 Sean X, Y y Z subconjuntos cerrados y no vacíos de $\mathbb{R}^n, \mathbb{R}^p$ y \mathbb{R}^q , respectivamente y sean T_1 y T_2 mapeos de punto a conjunto $T_1: X \rightarrow Y$, $T_2: Y \rightarrow Z$. El mapeo composición $T = T_2 \circ T_1$ es el mapeo de punto a conjunto $T: X \rightarrow Z$ definido por $T(x) = \bigcup \{T_2(y); y \in T_1(x)\}$. Frecuentemente se denota $T_2 \circ T_1$ por $T_2 T_1$.

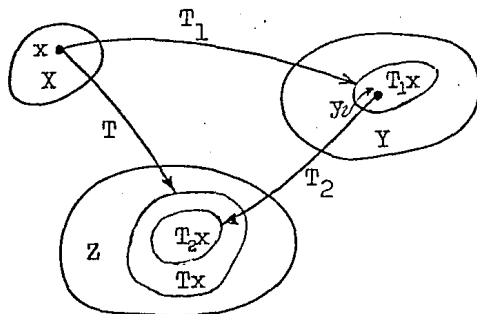


Figura 32. Composición de mapeos.

Teorema. 3.3.2 Sean X, Y y Z subconjuntos cerrados y no vacíos de $\mathbb{R}^n, \mathbb{R}^p$ y \mathbb{R}^q , respectivamente. Sean $T_1: X \rightarrow Y$ y $T_2: Y \rightarrow Z$ dos mapeos de punto a conjunto y considere el mapeo composición $T=T_2T_1$. Suponga que el mapeo T_1 es cerrado en el punto $x \in X$ y que el mapeo T_2 es cerrado en todo punto $y \in T_1(x)$. Suponga además que si $\{x_k\}_{k \in \mathbb{N}}$ es una sucesión infinita de elementos de X , tal que $x_k \rightarrow x$ y $y_k \in T_1(x_k)$, entonces existe una subsucesión convergente $\{y_{k_m}\}_{m \in \mathbb{N}}$ de la sucesión $\{y_k\}_{k \in \mathbb{N}}$. Entonces el mapeo composición T es cerrado en el punto x .

Prueba:

Sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión de elementos de X , tal que $x_k \rightarrow x$ cuando $k \rightarrow \infty$ y sea $\{z_k\}_{k \in \mathbb{N}}$ una sucesión que satisface $z_k \in T(x_k)$, tal que $z_k \rightarrow z \in Z$ cuando $k \rightarrow \infty$. Se verá que $z \in Tx$. Por definición del mapeo composición T , para cada $k \in \mathbb{N}$, existe un $y_k \in T_1(x_k)$, tal que $z_k \in T_2(y_k)$. Por hipótesis, existe una subsucesión convergente $\{y_{k_m}\}_{m \in \mathbb{N}}$ de la sucesión $\{y_k\}_{k \in \mathbb{N}}$, tal que $y_{k_m} \rightarrow y \in Y$. Puesto que T_1 es cerrado en $x \in X$ y como $x_{k_m} \rightarrow x$ y $y_{k_m} \rightarrow y$ cuando $m \rightarrow \infty$ con $y_{k_m} \in T_1(x_{k_m})$, ya que si $\{x_k\}_{k \in \mathbb{N}}$ converge a x , se tiene que $\{x_{k_m}\}_{m \in \mathbb{N}}$ también lo hace. Entonces $y \in T_1(x)$. Además T_2 es cerrado sobre $T_1(x)$ y como $y_{k_m} \rightarrow y$ y $z_{k_m} \rightarrow z$ cuando $m \rightarrow \infty$ con $z_{k_m} \in T_2(y_{k_m})$, entonces $z \in T_2(y)$. Así $z \in T_2(y) \in T_2(T_1(x)) = (T_2T_1)(x) = T(x)$.

Los dos resultados siguientes son consecuencia directa del teorema anterior.

Corolario. 3.3.3 Sean X, Y y Z subconjuntos cerrados y no vacíos de $\mathbb{R}^n, \mathbb{R}^p$ y \mathbb{R}^q , respectivamente. Sean $T_1: X \rightarrow Y$ y $T_2: Y \rightarrow Z$ dos mapeos de punto a conjunto. Suponga que el mapeo T_1 es cerrado en el punto $x \in X$ y que el mapeo T_2 es cerrado

en todo punto y $T_1(x)$. Si el conjunto Y es compacto. Entonces $T=T_2T_1$ es cerrado en $x \in X$.

Corolario. 3.3.4 Sean X, Y y Z subconjuntos cerrados y no vacíos de $\mathbb{R}^n, \mathbb{R}^p$ y \mathbb{R}^q , respectivamente. Sea $G: X \rightarrow Y$ una función continua y sea $T_2: Y \rightarrow Z$ un mapeo de punto a conjunto el cual es cerrado sobre $T_2(x)$. Entonces $T=T_2G$ es cerrado en el punto $x \in X$.

Note la importancia de la hipótesis de que existe una subsucesión convergente $\{y_{k_m}\}_{m \in \mathbb{N}}$ de elementos de Y . Sin esta hipótesis puede ser que aunque T_1 y T_2 sean ambos mapeos cerrados, el mapeo composición $T=T_2T_1$ no es necesariamente cerrado como se muestra en el siguiente ejemplo.

Ejemplo. 3.3.5 Sean $T_1, T_2: \mathbb{R} \rightarrow \mathbb{R}$ definidos por $T_1(x) = \begin{cases} \frac{1}{x} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$ y $T_2(y) = \{z; |z| \leq |y|\}$. Observe que T_1 y T_2 son mapeos cerrados sobre \mathbb{R} . Sea $T=T_2T_1$, entonces $T(x) = T_2(T_1(x)) = \{z; |z| \leq |T_1(x)|\}$, es decir, $T(x) = \{z; |z| \leq \left|\frac{1}{x}\right|\}$ si $x \neq 0$ y $T(0) = \{0\}$. Se verá que el mapeo T no es cerrado en el punto $x=0$.
Sea $\{x_k\}_{k \in \mathbb{N}}$ la sucesión definida por $x_k = \frac{1}{k}$ $k \in \mathbb{N}$. Note que $T(x_k) = \{z; |z| \leq k\}$ y por lo tanto $z_k = 1 \in T(x_k)$ para toda $k \in \mathbb{N}$. Sea $\{z_k\}_{k \in \mathbb{N}}$ la sucesión $\{1, 1, \dots, 1, \dots\}$, obviamente $z_k \rightarrow 1$ cuando $k \rightarrow \infty$ y sin embargo $1 \notin T(0) = \{0\}$. Por lo tanto, el mapeo T no es cerrado en el punto $x=0$ aún cuando T_1 y T_2 lo son. El teorema 3.3.2 no se aplica debido a que la sucesión $\{y_k\}_{k \in \mathbb{N}}$ con $y_k \in T_1(x_k) = \{k\}$ para $x_k = \frac{1}{k}$ no tiene subsucesiones convergentes.

Muchos algoritmos de programación no lineal, en cada iteración utilizan dos mapeos de punto a conjunto, uno de los cuales es cerrado y satisface las hipótesis del teorema 3.3.2 y otro que tiene como propósito hacer que el valor de alguna función de ascenso no disminuya. Se verá a continua-

ción que los algoritmos que poseen las propiedades antes mencionadas más cierta condición de compacidad sobre las sucesiones generadas, efectivamente convergen.

Teorema 3.3.6 Sea X un subconjunto cerrado y no vacío de \mathbb{R}^n y sea $\Omega \subset X$ un conjunto solución no vacío. Sea $F: \mathbb{R}^n \rightarrow \mathbb{R}$ una función continua y suponga que el mapeo de punto a conjunto $T_2: X \rightarrow X$ satisface la siguiente propiedad: Dado $x \in X$, $F(y) \geq F(x)$ para todo $y \in T_2(x)$. Sea $T_1: X \rightarrow X$ un mapeo de punto a conjunto que es cerrado sobre $X \setminus \Omega$, el cual satisface $F(y) > F(x)$ para toda $y \in T_1(x)$ cuando $x \notin \Omega$. Ahora considere el mapeo composición $T = T_2 \circ T_1$. Sea $x_1 \in X$ y suponga que la sucesión $\{x_k\}_{k \in \mathbb{N}}$ es generada como sigue: Si $x_k \in \Omega$, entonces el algoritmo se detiene; En caso contrario, tome $x_{k+1} \in T x_k$ reemplace k por $k+1$ e itere de nuevo. Suponga también que el conjunto $K = \{x; F(x) \geq F(x_1)\}$ es un subconjunto compacto de \mathbb{R}^n . Entonces el algoritmo converge sobre X , es decir, al algoritmo se detiene en un número finito de iteraciones con un punto en Ω , o bien todos los puntos de acumulación de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ pertenecen al conjunto solución Ω .

Prueba:

Si en alguna iteración $x_k \in \Omega$, entonces el algoritmo se detiene. Suponga ahora que una sucesión infinita es generada por el mapeo algorítmico T . Sea $\{x_{k_m}\}_{m \in \mathbb{N}}$ una subsucesión convergente a un punto $x \in X$, entonces por continuidad de la función F , se sigue que $F(x_{k_m}) \rightarrow F(x)$ cuando $m \rightarrow \infty$ y por el lema 3.2.1, se tiene que $\lim F(x_{k_m}) = F(x)$. Para probar que $x \in \Omega$, suponga que por el contrario $x \notin \Omega$ y considere la sucesión $\{x_{k_m+1}\}_{m \in \mathbb{N}}$. Por definición del mapeo T , note que $x_{k_m+1} \in T_2(y_{k_m})$, donde $y_{k_m} \in T_1(x_{k_m})$, teniéndose así que $y_{k_m}, x_{k_m+1} \in K$ para toda $m \in \mathbb{N}$, ya que, $F(x_1) \leq F(x_{k_m}) < F(y_{k_m}) \leq F(x_{k_m+1})$. Puesto que el conjunto

K es compacto, existen subsucesiones $\{y_{k_m p}\}_{p \in \mathbb{N}}$ y $\{x_{k_m p+1}\}_{p \in \mathbb{N}}$,
 tales que $y_{k_m p} \rightarrow y$ y $x_{k_m p+1} \rightarrow \bar{x}$ cuando $p \rightarrow \infty$ para algunos
 $y, \bar{x} \in X$. Por el lema 3.2.1 $\lim_{p \rightarrow \infty} F(x_{k_m p+1}) = F(\bar{x})$ y por lo tanto
 $F(\bar{x}) = F(x)$ y puesto que el mapeo T_1 es cerrado en $x \in \Omega$ y como
 $x_{k_m p} \rightarrow x$ y $y_{k_m p} \in T_1(x_{k_m p})$, entonces $y \in T_1(x)$ y por hipótesis
 $F(y) > F(x)$ y puesto que $x_{k_m p+1} \in T_2(y_{k_m p})$ para toda $p \in \mathbb{N}$, enton-
 ces también por hipótesis $F(x_{k_m p+1}) \geq F(y_{k_m p})$ para toda $p \in \mathbb{N}$,
 tomando el límite cuando $p \rightarrow \infty$, se tiene por continuidad de la
 función F que $F(\bar{x}) \geq F(y)$ y como ya se vió que $F(y) > F(x)$ se
 implica que $F(\bar{x}) > F(x)$, lo cual contradice el hecho de que
 $F(\bar{x}) = F(x)$. Por lo tanto $x \in \Omega$.

3.4 RAPIDEZ DE CONVERGENCIA

En el transcurso de la sección anterior se exploraron las condiciones para que un algoritmo converja. En esta sección será de interés desarrollar herramientas de análisis que permitan estudiar la rapidez con la que un algoritmo alcanza convergencia a un punto del conjunto solución.

A continuación se presentan dos definiciones que serán de utilidad para precisar la idea de rapidez de convergencia de un algoritmo.

Definición. 3.4.1 Sea $\{r_k\}_{k \in \mathbb{N}}$ una sucesión de números reales. Sea E el subconjunto de \mathbb{R} que consiste de aquellos números r , tales que existe una subsucesión $\{r_{k_m}\}_{m \in \mathbb{N}}$ con $r_{k_m} \rightarrow r$ cuando $m \rightarrow \infty$. Entonces

$$(i). \limsup_{k \rightarrow \infty} r_k = \sup E$$

$$(ii). \liminf_{k \rightarrow \infty} r_k = \inf E$$

Definición. 3.4.2 Sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión de elementos de \mathbb{R}^n que converge a un punto $x^* \in \mathbb{R}^n$. El orden de convergencia $\theta(\{x_k\}_{k \in \mathbb{N}})$ de la sucesión $\{x_k\}_{k \in \mathbb{N}}$ es el número

$$\theta(\{x_k\}_{k \in \mathbb{N}}) = \sup \left\{ p > 0; \limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} < \infty \right\}.$$

De la definición anterior se observa que si $\beta = \limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p}$, entonces para k suficientemente grande $\|x_{k+1} - x^*\| \approx \beta \|x_k - x^*\|^p$, con $\|x_k - x^*\| < 1$, teniéndose así que entre mayor es el valor de p más rápido es alcanzada la convergencia a x^*

Ejemplo. 3.4.3 Considere los mapeos algorítmicos $T, S: [0, \infty) \rightarrow [0, \infty)$ definidos por $Tx = \{\frac{1}{2}x\}$ y $Sx = \{\frac{1}{2}x^2\}$ y sea $\Omega = \{0\}$, entonces para el mapeo T y $x_1 > 0$, se tiene que

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} = \frac{|x_{k+1}|}{|x_k|^p} = \frac{\frac{1}{2} x_k}{x_k^p} = \frac{1}{2} x_k^{1-p}.$$

Para $p=1$, se tiene que

$$\limsup_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{1}{2},$$

para $p=2$

$$\limsup_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \limsup_{k \rightarrow \infty} \frac{1}{2} \cdot \frac{1}{x_k} = \infty,$$

ya que si $x_k \rightarrow 0$ cuando $k \rightarrow \infty$, entonces $\frac{1}{x_k} \rightarrow \infty$ cuando $k \rightarrow \infty$.

Teniéndose como consecuencia que $\theta_T(\{x_k\}_{k \in \mathbb{N}}) = 1$.

Para el mapeo algorítmico S y $x_1 > 0$, se tiene que

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} = \frac{x_k^2}{x_k^p}$$

Para $p=1$ y $p=2$ claramente $\limsup_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} < \infty$. Mientras que

para $p > 2$ se cumple que $\limsup_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} = \infty$. Por lo tanto

$\theta_S(\{x_k\}_{k \in \mathbb{N}}) = 2$. De donde S converge más rápido que T a Ω .

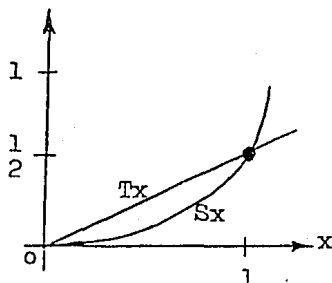


Figura 33. Mapeos algorítmicos del ejemplo 3.4.3

Existe una amplia clase de algoritmos de programación no lineal que poseen orden de convergencia uno. Se hará un breve análisis para este particular pero importante caso.

Definición. 3.4.4 Si una sucesión $\{x_k\}_{k \in \mathbb{N}}$ de elementos de \mathbb{R}^n converge a un punto $x^* \in \mathbb{R}^n$ de tal forma que

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \beta < 1.$$

Se dice entonces que la convergencia es lineal a razón β . Si $\beta = 0$, se dice que la convergencia es super-lineal.

De la definición anterior se observa que β es una medida útil en la determinación de cuál de dos o más algoritmos que convergen linealmente es mejor, siendo por supuesto, el de menor valor de β el que más rápido converge, ya que, para k suficientemente grande $\|x_{k+1} - x^*\| \approx \beta \|x_k - x^*\|$.

Ejemplo. 3.4.5 Considere la sucesión generada por el mapeo algorítmico T del ejemplo 3.4.3, entonces $\{x_k\}_{k \in \mathbb{N}}$ converge linealmente a razón $\beta = \frac{1}{2}$.

Ejemplo. 3.4.6

(i). Sea $x_k = \frac{1}{k}$ para toda $k \in \mathbb{N}$, entonces $x_k \rightarrow 0$ cuando $k \rightarrow \infty$ y

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{x_{k+1}}{x_k} = \frac{1}{1 + (1/k)} \rightarrow 1 \text{ cuando } k \rightarrow \infty.$$

Por lo tanto, la convergencia es lineal.

(ii). Sea $x_k = \left(\frac{1}{k}\right)^k$ para toda $k \in \mathbb{N}$, entonces $x_k \rightarrow 0$ cuando $k \rightarrow \infty$ y

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{x_{k+1}}{x_k} = \frac{1}{k(1 + (1/k))^{k+1}} \rightarrow 0 \text{ cuando } k \rightarrow \infty.$$

De donde la convergencia es superlineal.

Las definiciones anteriores de rapidez de convergencia son más bien de naturaleza global y no proporcionan información

sobre la rapidez con que el algoritmo progresa de una iteración a la siguiente. Una alternativa es tratar de encontrar una medida de rapidez promedio por paso después de cierto número de pasos.

Definición. 3.4.7 Sea $\{x_k\}_{k \in \mathbb{N}}$ una sucesión de elementos de \mathbb{R}^n convergente a un punto $x^* \in \mathbb{R}^n$. El orden promedio de convergencia $\bar{\theta}(\{x_k\}_{k \in \mathbb{N}})$ de la sucesión $\{x_k\}_{k \in \mathbb{N}}$, es el número

$$\bar{\theta}(\{x_k\}_{k \in \mathbb{N}}) = \inf \left\{ p > 1; \limsup_{k \rightarrow \infty} \|x_k - x^*\| \left(\frac{1}{p}\right)^k = 1 \right\}$$

Ejemplo. 3.4.8 Sea $a \in (0, 1)$ y sea $x_k = a^{2^k}$ para toda $k \in \mathbb{N}$, entonces $x_k \rightarrow 0$ cuando $k \rightarrow \infty$. Así

$|x_k - x^*| \left(\frac{1}{p}\right)^k = a^{2^k} / p^k$. Si $p=1$, entonces $\limsup_{k \rightarrow \infty} |x_k - x^*| \left(\frac{1}{p}\right)^k = 0$, si $p=2$, entonces $\limsup_{k \rightarrow \infty} |x_k - x^*| \left(\frac{1}{p}\right)^k = a$; pero si $p > 2$, entonces $\limsup_{k \rightarrow \infty} |x_k - x^*| \left(\frac{1}{p}\right)^k = 1$. Por lo tanto $\bar{\theta}(\{x_k\}_{k \in \mathbb{N}}) = 2$.

CAPITULO

4

EL ALGORITMO DE PAVIANI

CAPITULO 4. EL ALGORITMO DE PAVIANI

4.1 INTRODUCCION

En la mayoría de los casos cuando un algoritmo de programación no lineal para problemas restringidos, es implantado en una computadora digital, una considerable parte del tiempo de procesamiento es gastado para satisfacer requerimientos de factibilidad. D. Paviani, recientemente ha desarrollado un algoritmo descendente, más precisamente un algoritmo composición, el cual en cada iteración, utiliza la información cedida, tanto por puntos factibles como por ciertos puntos permitidos, no factibles, llamados puntos cuasi-factibles, de tal manera, que los límites de cuasi-factibilidad son gradualmente hechos más restrictivos en tanto que el mapeo algorítmico converge a un punto del conjunto solución, en donde, por supuesto, sólo puntos factibles son aceptados.

4.2 METODO DE BUSQUEDA POLIEDRICA

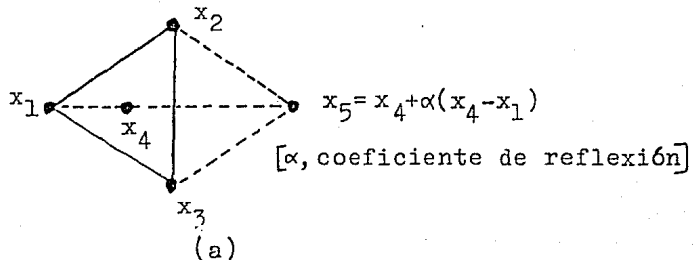
El método de búsqueda poliédrica que a continuación se describe, es usado con frecuencia para resolver problemas no restringidos, sin embargo, se verá en la siguiente sección, que este método tiene particular importancia en el desarrollo de algunas etapas del algoritmo de Paviani para problemas de programación no lineal restringidos.

El método de búsqueda poliédrica consiste, en construir en cada iteración un poliedro en \mathbb{R}^n y evaluar la función en cada uno de sus vértices. Después se considera, el vector que tiene como puntos inicial y final, el vértice del poliedro donde se obtuvo el valor máximo de la función obje-

tivo y el centroide del poliedro, respectivamente. A continuación se multiplica dicho vector un escalar constante positivo, llamado coeficiente de reflexión, el punto final del vector así obtenido es llamado punto de reflexión. Teniendo de esta manera, un nuevo poliedro que se obtiene del poliedro anterior, eliminando el vértice donde se obtuvo el valor máximo de la función objetivo e incluyendo el punto de reflexión. El procedimiento descrito es aplicado al nuevo poliedro y así sucesivamente.

Los poliedros obtenidos, en general, no tienen por que ser regulares. En cada iteración los lados del poliedro pueden ser expandidos o contraídos en la dirección definida por el vértice donde se obtuvo el valor máximo de la función objetivo y el centroide, dependiendo si el punto de reflexión ha producido un poliedro con un vértice de menor valor de la función objetivo que el mínimo de los valores de la función objetivo en los vértices del poliedro anterior, o bien si el punto de reflexión no produjo un poliedro con un vértice con valor menor de la función objetivo que el mínimo de los valores de la función objetivo en los vértices del poliedro anterior.

En caso de que se produzca un nuevo poliedro mediante la contracción del punto de reflexión y no se obtenga un vértice con valor de la función objetivo menor que el mínimo de los valores de la función objetivo en los vértices del poliedro anterior, entonces todos los lados del poliedro anterior se reducen por mitad quedando fijo el vértice donde se obtuvo el valor mínimo de la función objetivo.



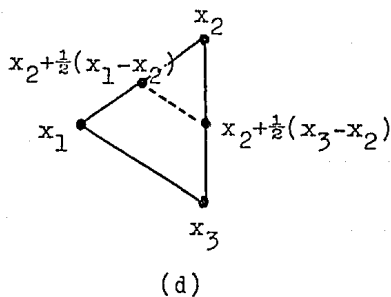
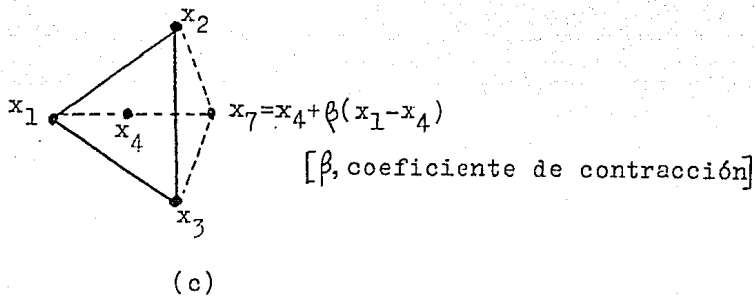
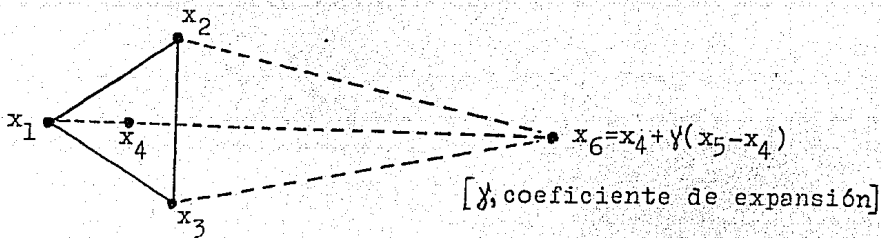


Figura 34. Suponga que en los vértices x_1 y x_2 se obtuvieron los valores máximo y mínimo de la función objetivo, respectivamente, y que x_4 es el centroide del poliedro con vértices x_1, x_2 y x_3 ; (a) muestra una reflexión; (b) una expansión; (c) una contracción y; (d) una reducción.

Lo anterior es formalizado a continuación:

Sea $P = \{x_i^k\}_{1 \leq i \leq n+1}$, $x_i^k \in \mathbb{R}^n$, el conjunto de vértices del poliedro de búsqueda en la k -ésima iteración y sean

$$y^* = \max_{1 \leq i \leq n+1} \{f(x_i^k)\} \quad \text{y} \quad z^* = \min_{1 \leq i \leq n+1} \{f(x_i^k)\}$$

por comodidad se denotarán $x_{y^*}^k$ y $x_{z^*}^k$ los vértices del poliedro donde se alcanzan los valores y^* y z^* , respectivamente. Las coordenadas del centroide, el cual se denotará por x_{n+2}^k , están dadas por

$$x_{n+2, j}^k = \frac{1}{n+1} \sum_{i=1}^{n+1} x_{i, j}^k, \quad \text{donde } x_i^k = (x_{i,1}^k, x_{i,2}^k, \dots, x_{i,n}^k)^t.$$

Los procesos de reflexión, expansión, contracción y reducción se precisan como sigue:

(i). Reflexión

El punto de reflexión x_{n+3}^k es obtenido mediante la relación

$$x_{n+3}^k = x_{n+2}^k + \alpha(x_{n+2}^k - x_{y^*}^k),$$

donde el número $\alpha > 0$ es el coeficiente de reflexión.

(ii). Expansión

Si $f(x_{n+3}^k) < z^*$, entonces determine el punto de expansión mediante la expresión

$$x_{n+4}^k = x_{n+2}^k + \gamma(x_{n+3}^k - x_{n+2}^k),$$

donde $\gamma > 1$ es el coeficiente de expansión. Reemplace $x_{y^*}^k$ por x_{n+4}^k .

(iii). Contracción

Si $f(x_{n+3}^k) > f(x_1^k)$ para toda i diferente del índice donde se obtuvo y^* , calcule el punto de contracción como

$$x_{n+5}^k = x_{n+2}^k + \beta(x_{y^*}^k - x_{n+2}^k),$$

donde $0 < \beta < 1$ es el coeficiente de contracción. Reemplace $x_{y^*}^k$ por x_{n+5}^k .

(iv). Reducción

Si $f(x_{n+1}^k) > y^*$, entonces reduzca todos los vectores $x_1^k - x_{z^*}^k$ para $i=1, 2, \dots, n+1$, mediante la expresión

$$x_i^k = x_{z^*}^k + \frac{1}{2}(x_i^k - x_{z^*}^k) \quad i=1, 2, \dots, n+1 \text{ y continúe la búsqueda.}$$

El criterio para terminar la búsqueda es que se satisfaga

$$\left\{ \frac{1}{n+1} \sum_{i=1}^{n+1} (f(x_i^k) - f(x_{n+2}^k))^2 \right\}^{\frac{1}{2}} \leq \epsilon,$$

para un $\epsilon > 0$ predeterminado.

Los vértices del poliedro inicial pueden ser determinados mediante la relación

$$x_i^0 = x^0 + D_i \quad i=1, 2, \dots, n+1,$$

donde x^0 es un vértice inicial proporcionado y los vectores D_i son los vectores columna en \mathbb{R}^n de la matriz de $n \times (n+1)$

$$D = \begin{bmatrix} 0 & d_1 & d_2 & d_2 & \cdot & \cdot & \cdot & d_2 \\ 0 & d_2 & d_1 & d_2 & \cdot & \cdot & \cdot & d_2 \\ 0 & d_2 & d_2 & d_1 & \cdot & \cdot & \cdot & d_2 \\ \vdots & \vdots & \vdots & \vdots & \cdot & \cdot & \cdot & \vdots \\ 0 & d_2 & d_2 & d_2 & \cdot & \cdot & \cdot & d_1 \end{bmatrix},$$

en donde

$$d_1 = \frac{t}{n\sqrt{2}}(\sqrt{n+1} + n - 1) \quad \text{y} \quad d_2 = \frac{t}{n\sqrt{2}}(\sqrt{n+1} - 1),$$

siendo t la distancia entre dos vértices.

Ejemplo. 4.2.1 El procedimiento descrito para determinar los vértices del poliedro inicial con el vértice fijo $x^0=0$, $n=2$ y $t=1$ proporciona los puntos:

$$x_1^0 = 0, \quad x_2^0 = (0.965, 0.259)^t \quad \text{y} \quad x_3^0 = (0.259, 0.965)^t.$$

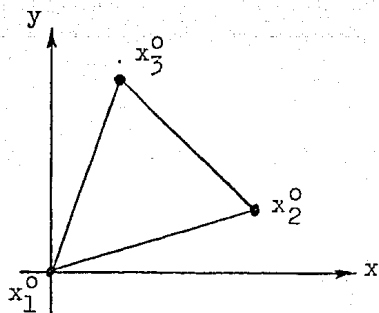
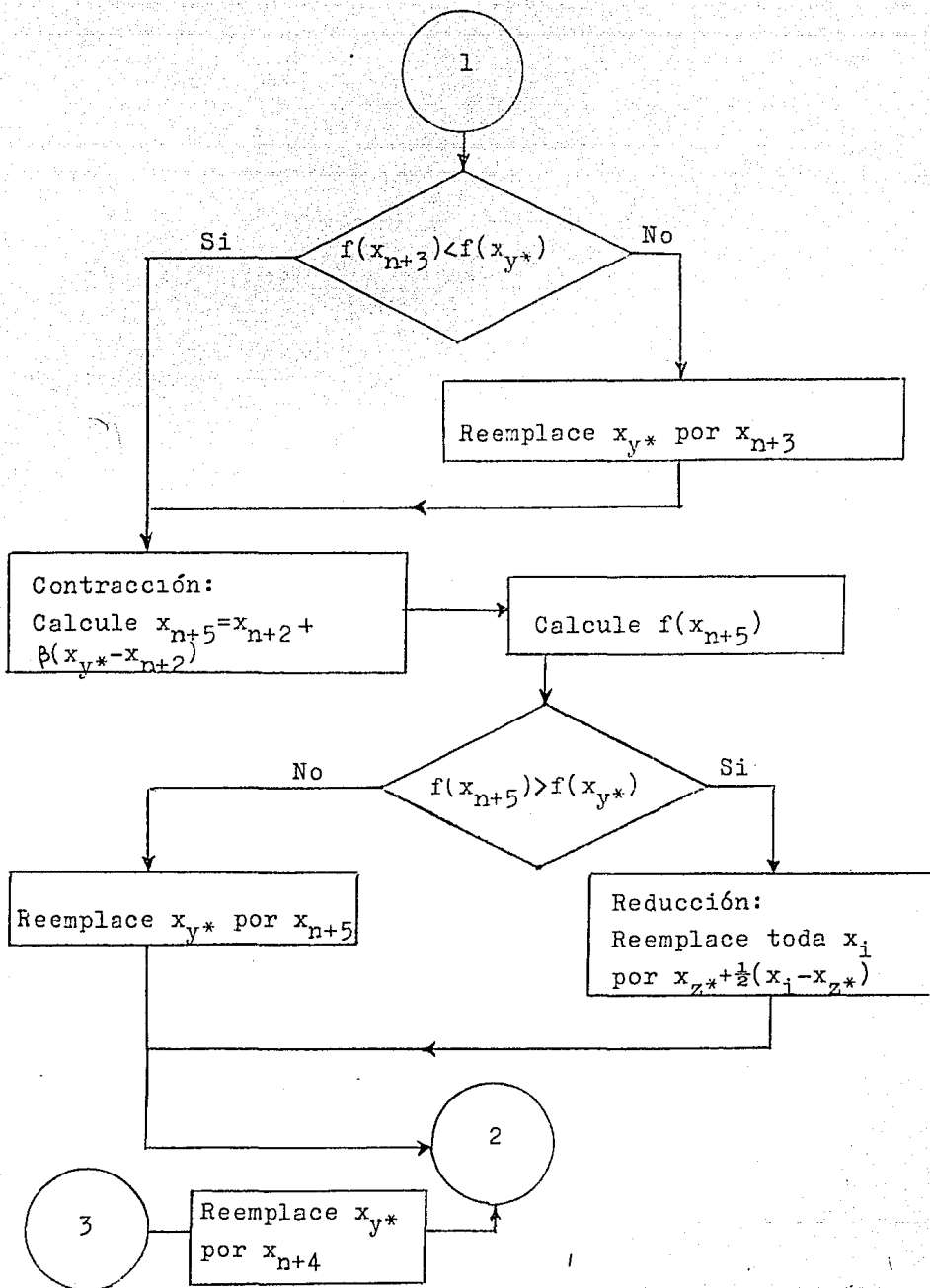


Figura 35. Pólhedro del ejemplo 4.2.1

(Cont.)



4.3 DESCRIPCIÓN DEL ALGORITMO DE PAVIANI

El problema general de programación no lineal es repetido aquí por conveniencia

(I). Minimizar $f(x)$

$$\begin{aligned} \text{sujeto a } g_i(x) &\leq 0, \quad i=1,2,\dots,m \\ h_i(x) &= 0, \quad i=m+1,\dots,p \\ x &\in X, \end{aligned}$$

donde $X \subset \mathbb{R}^n$, $X \neq \emptyset$ y las funciones $f, g_i, h_i: \mathbb{R}^n \rightarrow \mathbb{R}$ pueden ser lineales y/o no lineales.

Se comenzará por replantear el problema (I) de la siguiente manera:

(II). Minimizar $f(x)$

$$\text{sujeto a } \Phi^k - S(x) \geq 0,$$

donde S es una funcional positiva de las restricciones y Φ^k es el valor del criterio de tolerancia regulable para violar factibilidad durante la k -ésima iteración. Dicho criterio es definido por:

$$\Phi^k = \min \left\{ \Phi^{k-1}, \frac{1}{n+1} \sum_{i=1}^{n+1} \|x_i^k - x_{n+2}^k\| \right\}, \quad \Phi^0 = t,$$

donde

t = distancia entre dos vértices del poliedro inicial de búsqueda.

x_i^k = i -ésimo vértice en \mathbb{R}^n de poliedro de búsqueda $1 \leq i \leq n+1$, en la k -ésima iteración.

x_{n+2}^k = centroide del poliedro de búsqueda en la k -ésima iteración.

La funcional S de las restricciones es definida mediante

$$S(x) = + \left\{ \sum_{i=1}^m U_i g_i^2(x) + \sum_{i=m+1}^p h_i^2(x) \right\}^{\frac{1}{2}}$$

donde U_i es el operador de Heaviside,

$$U_i = \begin{cases} 0, & \text{si } g_i \leq 0 \\ 1, & \text{si } g_i > 0 \end{cases}$$

Observe que $S(x) \geq 0$ para toda $x \in \mathbb{R}^n$ y que $\frac{1}{n+1} \sum_{i=1}^{n+1} \|x_i^k - x_{n+2}^k\|$ es la distancia promedio de cada vértice al centroide.

El rango permitido para violar factibilidad es determinado a continuación. Como $S(x) \geq 0$ para todo $x \in \mathbb{R}^n$, entonces cualquier vector $x \in \mathbb{R}^n$ satisface una y sólo una de las siguientes tres condiciones cuando $\Phi^k > 0$:

- (i). $S(x) = 0$
- (ii). $0 < S(x) \leq \Phi^k$
- (iii). $S(x) > \Phi^k$

En el caso en que un punto $x \in \mathbb{R}^n$ satisface (i), se tiene entonces que x es factible. Si x satisface la condición (ii) se dirá que x es un punto cuasi-factible. Por último si x satisface (iii), se dirá que x es no factible.

La región de cuasifactibilidad es definida mediante $\Phi^k - S(x) \geq 0$.

Para ver que la solución del problema (I) coincide con la del problema (II) cuando ésta existe, basta observar que $0 \leq \Phi^k \leq \Phi^{k-1}$ con $\lim_{k \rightarrow \infty} \Phi^k = 0$, es decir, en el límite la distancia promedio entre los vértices del poliedro y el centroide es cero, lo cual equivale a que el poliedro de búsqueda se ha contraído a un sólo punto, teniéndose como consecuencia que ya no es posible mejorar el valor de la función objetivo. Además como $0 \leq S(x) \leq \Phi^k$, en el límite se tendrá que $S(x) = 0$, obteniéndose así un punto factible.

A continuación se presentan los pasos del algoritmo de Paviani.

Proporcione los parámetros del método de búsqueda poliédrica:

α = factor de reflexión,

β = factor de contracción,

γ = factor de expansión,

t = distancia entre los vertices del poliedro inicial,

x^0 = vértice inicial propuesto

Proporcione además un criterio de tolerancia para cuasi-factibilidad $\epsilon > 0$, para detener el algoritmo de Paviani.

Paso 0. Obtenga los vértices del poliedro inicial $\{x^0, x_2^0, \dots, x_{n+1}^0\}$.

Paso 1. Asigne $k=0$

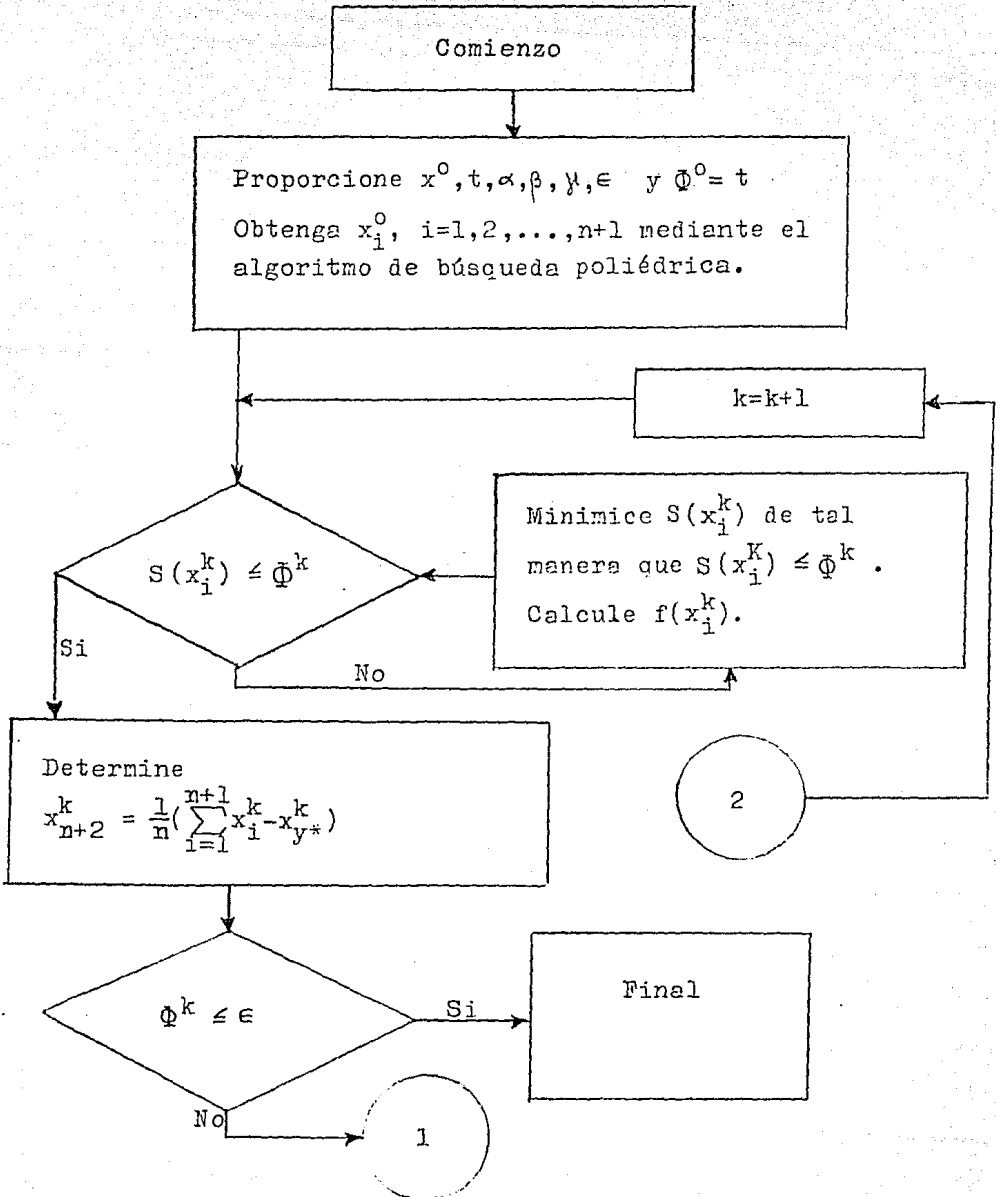
Paso 2. Calcule $x_{y^*}^k$ y $x_{z^*}^k$ y utilice el método de búsqueda poliédrica para mejorar el valor de la función objetivo. Obteniéndose así un nuevo poliedro con vértices $\{x^k, x_2^k, \dots, x_{n+1}^k\}$. Si $\Phi^k - S(x) \geq 0$ continúe con el paso 3; En caso contrario ir al paso 4.

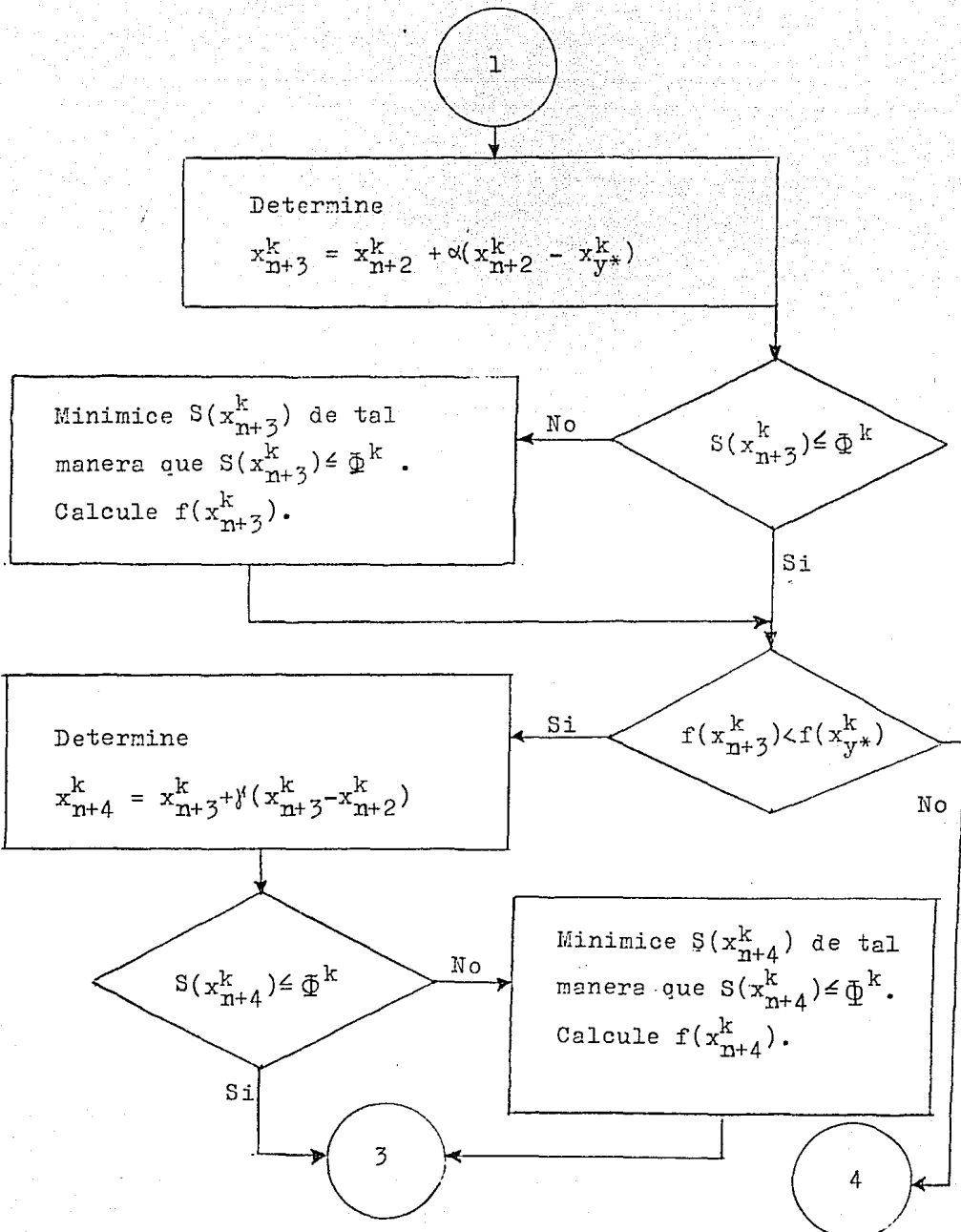
Paso 3. Si $\Phi^k > \epsilon$, asigne $k=k+1$ y regrese al paso 2. En caso contrario ir al paso 5.

Paso 4. Minimice $S(x)$ mediante búsqueda poliédrica hasta que $S(x) \leq \Phi^k$, haga $x=x^k$ y regrese al paso 2.

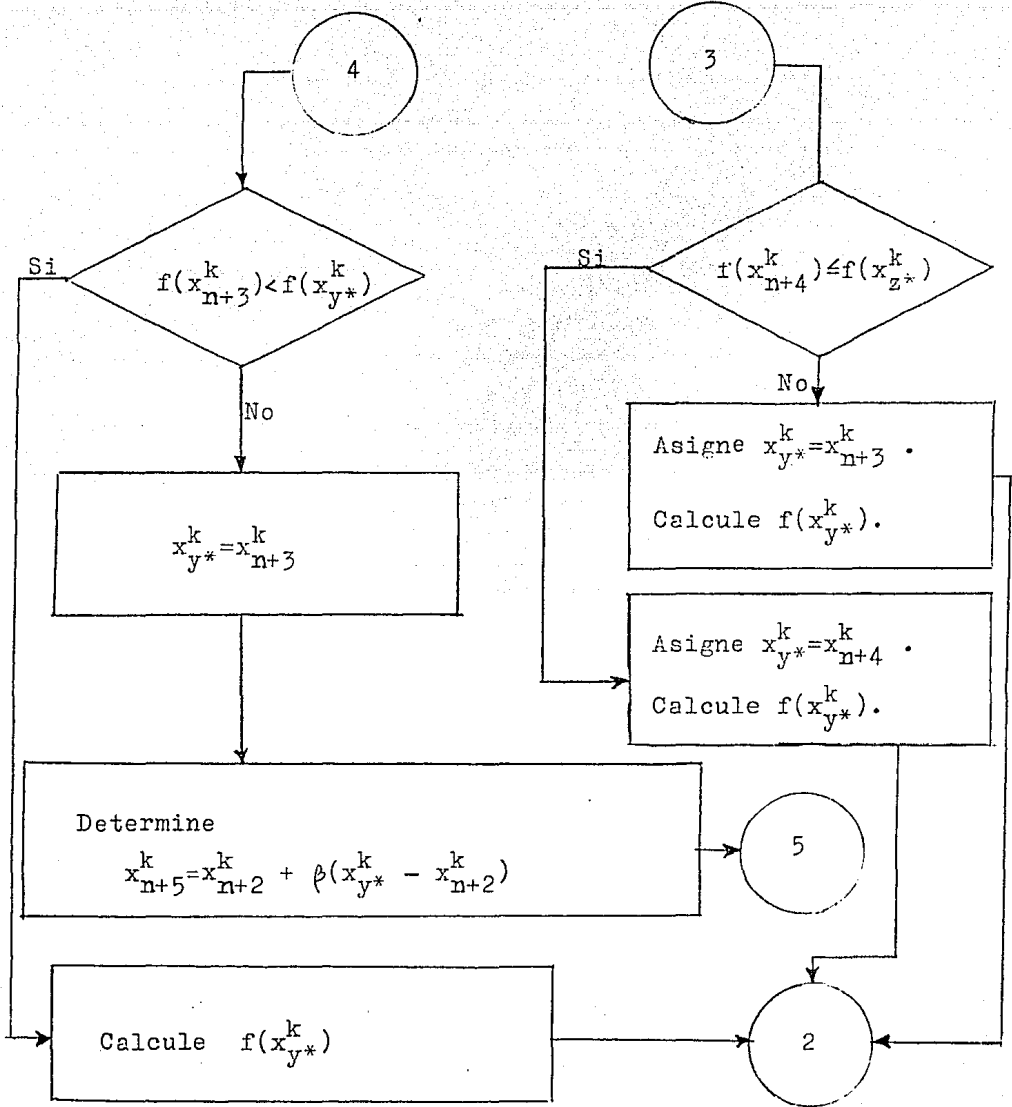
Paso 5. El algoritmo es detenido.

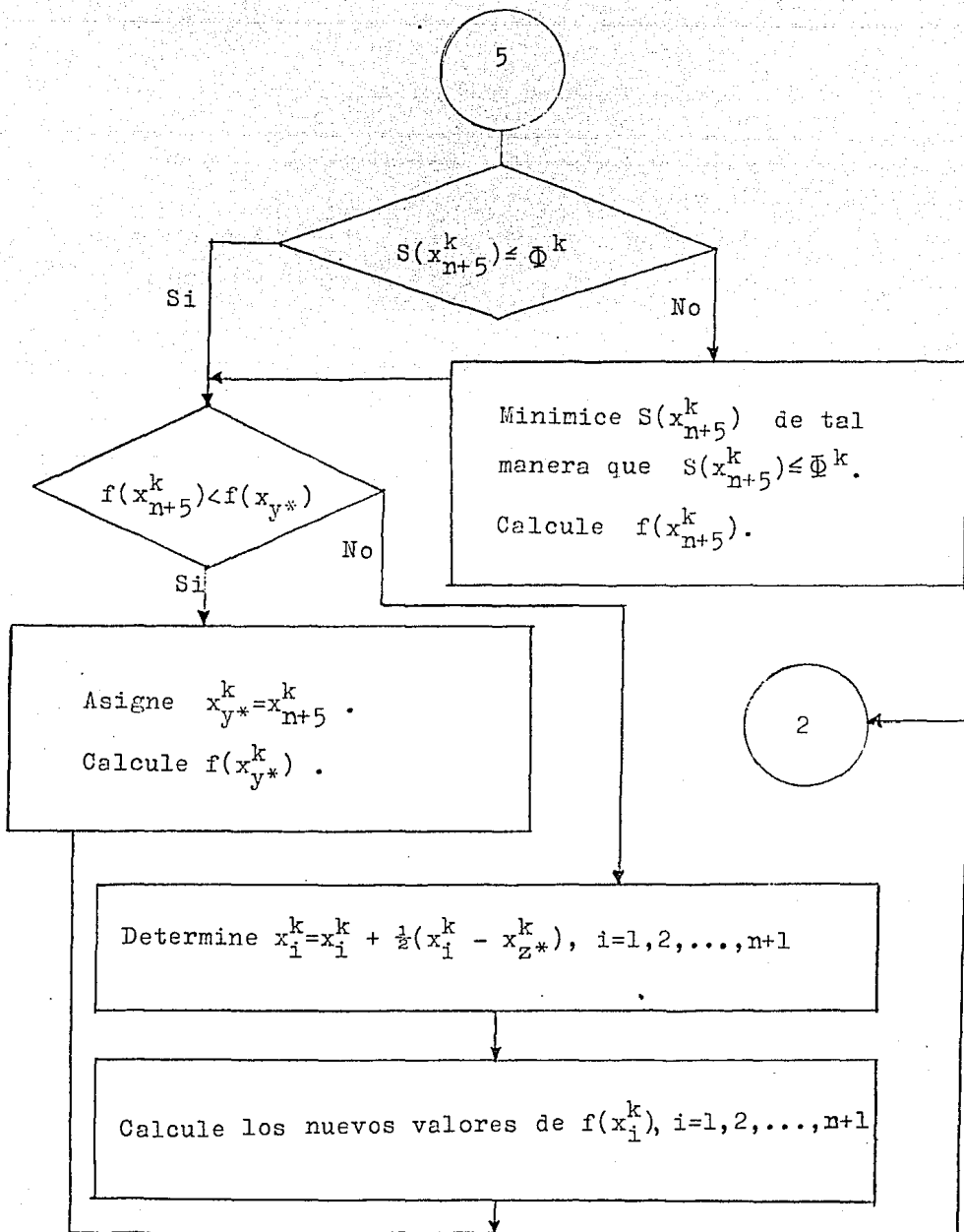
Diagrama de flujo del algoritmo de Paviani.





(Cont.)





4.4 ESTRUCTURA DEL PROGRAMA ALGPAV

El programa ALGPAV consiste de un programa principal y cuatro subrutinas en lenguaje Fortran IV.

El programa principal controla la ejecución de las subrutinas y consiste de dos partes. En la primera, se leen los datos requeridos: título, número total de variables independientes, número total de restricciones que son igualdades, número total de restricciones que son desigualdades, distancia entre los vértices del poliedro inicial, criterio de convergencia para detener el algoritmo y el vértice propuesto para el poliedro inicial. En esta misma parte, se asignan valores a los coeficientes de reflexión, contracción y expansión, y se inicializan contadores de iteración y de impresión de información. En la segunda parte, mediante búsqueda poliédrica se mejora el valor de la función objetivo. En caso de que en alguna iteración el vértice para el cual se obtiene el mínimo valor de la función objetivo no sea un punto cuasi-factible, se hace un llamado alternativo de las subrutinas FACTIB y SUMR. En la subrutina FACTIB se minimiza la suma de los cuadrados de las restricciones violadas y en la subrutina SUMR se calculan los valores de los cuadrados de las restricciones violadas. En la subrutina COMENZ se actualizan las distancias entre los vértices del poliedro generado.

Mediante la subrutina ESCRIB se imprime para cada iteración el valor de la función objetivo, el vector obtenido y los valores de las restricciones y por último mediante la subrutina PROBLE se introduce al programa la función objetivo y las restricciones del problema.

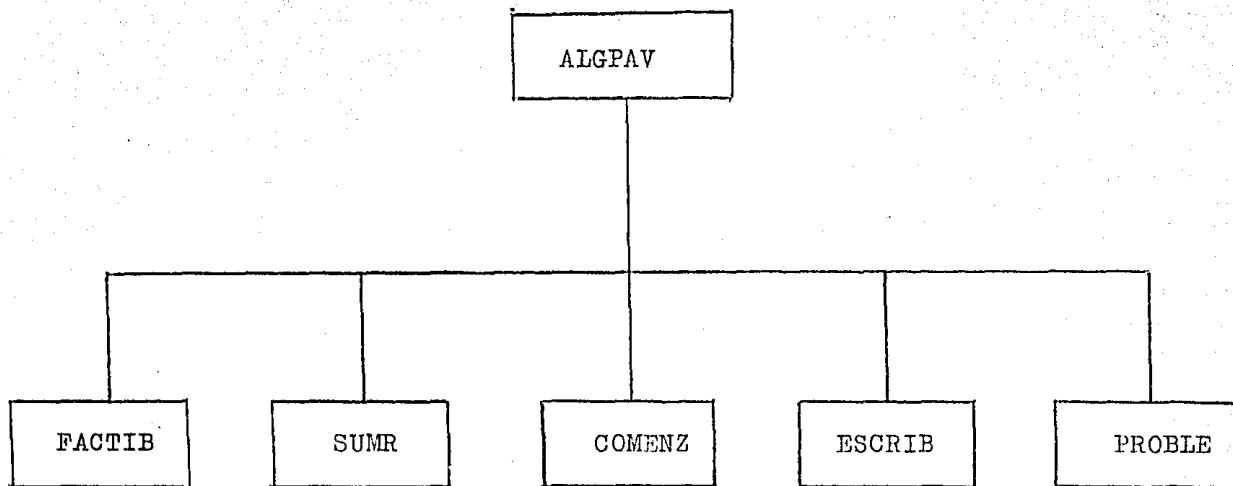


Figura 36. Estructura del programa ALGPAV.

4.5 ENTRADA DE DATOS DEL PROGRAMA ALGPAV

Tarjeta tipo 1

Columnas	Parámetros	Formato
1-80	TITULO, título del problema.	80A1

Tarjeta tipo 2

Columnas	Parámetros	Formato
1-5	NVI, número de variables independientes	I5
6-10	NRI, número de restricciones que son igualdades	I5
11-15	NRD, número de restricciones que son desigualdades	I5
16-25	TAM, distancia entre vértices del poliedro inicial	F10.5
26-32	EPS, criterio de convergencia para detener el algoritmo	F7.5
33-35	NUMITE, número máximo de iteraciones permitidas	I3

Tarjetas tipo 3

Columnas	Coordenadas del vértice inicial propuesto	Formato
1-10	X(1), primera coordenada del vértice inicial propuesto	F10.5
10-20	X(2), segunda coordenada del vértice inicial propuesto	F10.5
.	.	.
.	.	.
.	.	.

Se deben usar tantas tarjetas del tipo 3 como sean necesarias para leer todas las coordenadas del vértice inicial propuesto.

4.6 USO DEL PROGRAMA ALGPAV

Para ilustrar el acceso al programa ALGPAV, se resolverá mediante dicho programa el siguiente problema de programación no lineal restringida.

$$\begin{aligned} \text{Minimizar } f(x) &= \|x\|^2, \quad x \in \mathbb{R}^2 \\ \text{sujeto a } g_1(x) &= x_1 \geq 0 \\ g_2(x) &= x_2 \geq 0 \\ g_3(x) &= \|x\|^2 - 9x_2 + 4.25 = 0 \end{aligned}$$

En primer lugar el programa ALGPAV necesita que las restricciones que son desigualdades sean puestas en la forma $g_i(x) \geq 0$ $i=1,2,\dots,m$.

Dentro del programa ALGPAV se hacen las asignaciones $\alpha=1$, $\beta=0.5$ y $\gamma=2$, aunque el programa está diseñado, de tal manera que el usuario asigne los valores que desee.

El vértice inicial que se propone es $(4,4.5)^t$, la distancia entre los vértices del poliedro inicial se tomará como $t=1$. El criterio de tolerancia para detener el algoritmo será $\epsilon=0.00001$ y el número máximo permitido de iteraciones será de 40.

Claramente

NVI=2 (número de variables independientes)

NRI=1 (número de restricciones que son igualdades)

NRD=2 (número de restricciones que son desigualdades)

El problema planteado es introducido al programa ALGPAV a través de la subrutina PROBLE como se muestra a continuación:

```
SUBROUTINE PROBLE(INO)
  DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),
  ISR(50),RALD(100)
  COMMON/I/NVI,NRI,NRD,PASO,ALPHA,BETA,GAMMA,IN,INF,FDIFER,SECL,KC1,
  2KC2,KC3,KC4,KC5,KC6,KC7,KC8,KC9,X,X1,X2,R,SUM,F,SR,RALD,
  3CRUCE
```

```

COMMON/2/LFEAS,M5,M6,M7,M8,M9,S1A,S2A,S3A,TAN
IF(INQ.EQ.3) GO TO 3
IF(INQ.EQ.2) GO TO 2
C
C *****
C ** RESTRICCIONES QUE SON IGUALDADES **
C *****
C
  R(1) = X(1)**2 + X(2)**2 - 9.*X(2) + 4.25
  GO TO 5
C
C *****
C ** RESTRICCIONES QUE SON DESIGUALDADES **
C *****
C
  2 R(2) = X(1)
  R(3) = X(2)
  GO TO 5
C
C *****
C ** FUNCION OBJETIVO **
C *****
C
  3 R(4) = X(1)**2 + X(2)**2
  5 RETURN
  FND

```

La entrada de datos es hecha de acuerdo a la sección anterior, obteniéndose así el siguiente conjunto de datos

```

***** EJEMPLO DE TESIS *****
2 1 2 1.0.000140
3. 4.5

```

Los resultados obtenidos con el programa ALGPAV son mostrados en las siguientes hojas. Observe también que

	x_1	x_2
Solución exacta	0.0	0.5
Solución por programa	0.0140	0.5000

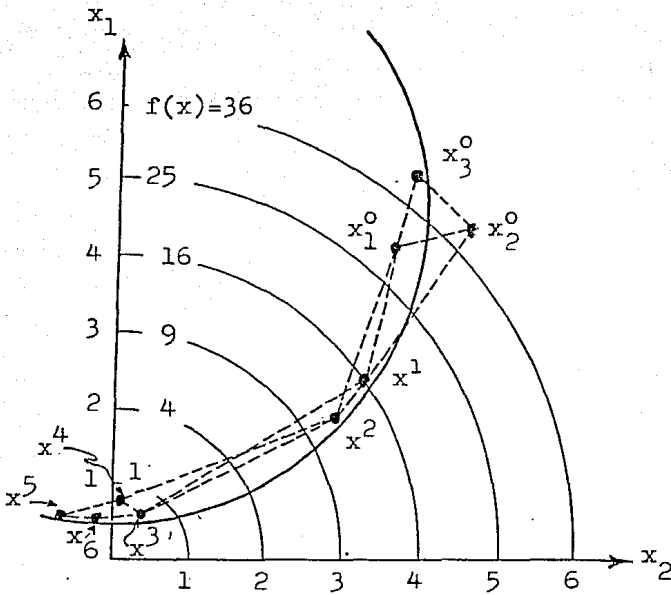


Figura 37. Evolución del programa ALGPAV en el problema planteado

NUMERO DE ETAPA DE CALCULOS = 1

CRITERIO DE TOLERANCIA ACTUAL= .400000E+01

VALOR DE LA FUNCION OBJETIVO = .2436572E+00

LOS VALORES DE LOS VECTORES INDEPENDIENTES SON

.1384562E-01 .4934222E+00

LOS VALORES DE LAS RESTRICCIONES QUE SON IGUALDADES SON

.5285744E-01

LOS VALORES DE LAS RESTRICCIONES QUE SON DESIGUALDADES SON
.1384562E-01 .4934222E+00

0

NUMERO DE ETAPA DE CALCULOS = 12

CRITERIO DE TOLERANCIA ACTUAL= .138291E+00

VALOR DE LA FUNCION OBJETIVO = .2562145E+00

LOS VALORES DE LOS VECTORES INDEPENDIENTES SON

.1403069E-01 .5000176E+00

LOS VALORES DE LAS RESTRICCIONES QUE SON IGUALDADES SON

.5592247E-04

LOS VALORES DE LAS RESTRICCIONES QUE SON DESIGUALDADES SON
.1403069E-01 .5000176E+00

0

NUMERO DE ETAPA DE CALCULOS = 24

CRITERIO DE TOLERANCIA ACTUAL= .138291E+00

VALOR DE LA FUNCION OBJETIVO = .2502215E+00

LOS VALORES DE LOS VECTORES INDEPENDIENTES SON

.1403089E-01 .5000246E+00

LOS VALORES DE LAS RESTRICCIONES QUE SON IGUALDADES SON

.7392487E-10

LOS VALORES DE LAS RESTRICCIONES QUE SON DESIGUALDADES SON

.1403089E-01 .5000246E+00

0

NUMERO DE ETAPA DE CALCULOS = 36

CRITERIO DE TOLERANCIA ACTUAL = .138291E+00

VALOR DE LA FUNCION OBJETIVO = .2502215E+00

LOS VALORES DE LOS VECTORES INDEPENDIENTES SON

.1403089E-01 .5000246E+00

LOS VALORES DE LAS RESTRICCIONES QUE SON IGUALDADES SON

0.

LOS VALORES DE LAS RESTRICCIONES QUE SON DESIGUALDADES SON

.1403089E-01 .5000246E+00

0

4.7 PROGRAMA FORTRAN ALGPV

```

C ***** DESCRIPCION DE VARIABLES Y PARAMETROS *****
C **
C ** NVI   NUMERO TOTAL DE VARIARLES INDEPENDIENTES
C ** NRI   NUMERO TOTAL DE RESTRICCIONES QUE CORRESPONDEN A
C **      IGUALDADES.
C ** NRD   NUMERO TOTAL DE RESTRICCIONES QUE CORRESPONDEN A
C **      DESIGUALDADES.
C ** TAM   LONGITUD DEL BORDE DEL POLIEDRO INICIAL
C ** EPS   CRITERIO DE CONVERGENCIA PARA TERMINAR LA BUSQUEDA
C ** ALPHA COEFICIENTE DE REFLEXION
C ** BETA  COEFICIENTE DE CONTRACCION
C ** GAMMA COEFICIENTE DE EXPANSION
C ** X(I)  VECTOR PROPUESTO PARA INICIAR LA BUSQUEDA
C ** FDIFER EL CRITERIO DE TOLERANCIA PARA VIULAR RESTRICCIONES
C ** NCONT CONTADOR PARA IMPRIMIR LA INFORMACION DE CADA
C **      (NVI+1)-FASE
C ** INFMIN INDICE PARA IDENTIFICAR LA INFORMACION RELACIONADA CON
C **      EL VALOR MINIMO DE LA FUNCION OBJETIVO EN EL POLIEDRO
C **      MAS RECIENTE
C ** INFMAX INDICE PARA IDENTIFICAR LA INFORMACION RELACIONADA
C **      CON EL VALOR MAXIMO DE LA FUNCION OBJETIVO EN EL
C **      POLIEDRO MAS RECIENTE
C ** ICONT CONTADOR PARA REGISTRAR EL NUMERO DE ITERACIONES
C **      EN EL POLIEDRO MAS RECIENTE
C ** LSEG  INDICE PARA IDENTIFICAR LA INFORMACION RELACIONADA
C **      CON EL VALOR MAS PROXIMO AL VALOR MAXIMO DE LA FUNCION
C **      OBJETIVO.
C *****
C
C DIMENSION TITULO(80)
C DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),
C 1SR(50),RALD(100)
C COMMON/1/NVI,NRI,NRD,PASO,ALPHA,BETA,GAMMA,IN,INF,FDIFER,SECL,KC1,
C 1KC2,KC3,KC4,KC5,KC6,KC7,KC8,KC9,X,X1,X2,R,SUM,F,SR,RALD,
C 2CRUCE
C COMMON/2/LFEAS,M5,M6,M7,M8,M9,S1A,S2A,S3A,TAM
C
C *****
C ** TITULO DEL PROBLEMA **
C *****
C
C READ(5,759) (TITULO(I),I=1,80)
C
C *****
C ** PARAMETROS DEL PROBLEMA **
C *****
C
C READ(5,1) NVI,NRI,NRD,TAM,EPS,NUMITE
C ALPHA = 1.
C BETA = 0.5
C GAMMA = 2.
C 10 CALL SECOND(TIEMPO)
C PASO=TAM

```

```

C *****
C ** VECTOR PROPUESTO PARA INICIAR LA BÚSQUEDA **
C *****
C
C   READ(5,2) (X(I), I=1,NVI)
C   WRITE(6,106)
C   WRITE(6,759) (TITULO(I), I=1,R0)
C   WRITE(6,756) NVI,NRI,NRD,TAM,EPS,TIEMPO
C
C *****
C ** PARAMETROS AUXILIARES **
C *****
C
C   KC1 = NVI + 1
C   KC2 = NVI + 2
C   KC3 = NVI + 3
C   KC4 = NVI + 4
C   KC5 = NVI + 5
C   KC6 = NRI + NRD
C   KC7 = NRI + 1
C   KC8 = NRI + NRD
C   KC9 = KC8 + 1
C   M = NVI - NRI
C   N1 = M + 1
C   IF(N1.GE.3) GO TO 50
C   N1 = 3
C   M = 2
50  N2 = M + 2
C   N3 = M + 3
C   N4 = M + 4
C   N5 = M + 5
C   N6 = M + 6
C   N7 = M + 7
C   N8 = M + 8
C   XN = M
C   XNX = NVI
C   XN1 = N1
C   S1A = 0.5*(SORT(5.) - 1.)
C   S2A = S1A**2
C   S3A = S1A*S2A
C   M5 = NVI + 5
C   M6 = NVI + 6
C   M7 = NVI + 7
C   M8 = NVI + 8
C   M9 = NVI + 9
C
C *****
C ** ACTUALIZACION DE CONTADORES **
C *****
C
C   ICONT = 1
C   NCONT = 1
C   WRITE(6,115)
C   WRITE(6,116) (X(J), J=1,NVI)
C
C *****
C ** CRITERIO DE TOLERANCIA PARA VIOLAR RESTRICCIONES **
C *****
C
C   FDIFER = 2.*(NRI+1)*PASO
C   CRUCE = FDIFER
C   IN = M1
C   CALL SUMR
C   SR(N1) = SORT(SECL)

```

```

WRITE(6,763) FDIFER,SR(N1)
IF(SR(N1).LT.FDIFER) GO TO 341
CALL ESCRIB
WRITE(6,757)
INF = N1
PASO = 0.05*FDIFER
CALL FACIIB
WRITE(6,764)
WRITE(6,116) (X2(INF,J),J=1,NVI)
WRITE(6,765) SR(INF)
IF(CRUCE.LT.1.0E-09) GO TO 80
341 WRITE(6,35)
WRITE(6,758) ICONT,FDIFER
CALL ESCRIB
FTER = R(KC9)

```

```

C
C *****
C ** CALCULO DEL CENTROIDE DE TODOS LOS VERTICES DEL POLIEDRO INICIAL**
C *****
C

```

```

PASO1 = PASO*(SQRT(XNX + 1.) + XNX - 1.)/(XNX*SQRT(2.))
PASO2 = PASO*(SQRT(XNX + 1.) - 1.)/(XNX*SQRT(2.))
ETA = (PASO1 + (XNX - 1.)*PASO2)/(XNX + 1.)
DO 4 J= 1,NVI
X(J) = X(J) - ETA
4 CONTINUE
CALL COMENZ
DO 9 I=1,N1
DO 9 J=1,NVI
X2(I,J) = X1(I,J)
9 CONTINUE
DO 5 I=1,N1
IN = I
DO 6 J=1,NVI
6 X(J) = X2(I,J)
CALL SUMR
SR(I) = SQRT(SECL)
IF(SR(I).LT.FDIFER) GO TO 8
CALL FACIIB
IF(CRUCE.LT.1.0E-09) GO TO 80
8 CALL PROBLE(3)
F(I)= R(KC9)
5 CONTINUE
1000 PASO=0.05*FDIFER
ICONT = ICONT + 1

```

```

C
C *****
C ** SELECCION DEL MAXIMO VALOR DE LA FUNCION OBJETIVO **
C ** A PARTIR DE LOS VERTICES DEL POLIEDRO **
C *****

```

```

FH = F(1)
INFMAX = 1
DO 16 I=2,N1
IF(F(I).LT.FH) GO TO 16
FH = F(I)
INFMAX = I
16 CONTINUE

```

```

C
C *****
C ** SELECCION DEL MINIMO VALOR DE LA FUNCION OBJETIVO **
C ** A PARTIR DE LOS VERTICES DEL POLIEDRO **
C *****

```

```

41 FL = F(1)

```

```

INFMIN = 1
DO 17 I=2,N1
IF (FL.LT.F(I)) GO TO 17
FL = F(I)
INFMIN = I
17 CONTINUE
DO 86 J=1,NVI
86 X(J) = X2(INFMIN,J)
IN = INFMIN
CALL SUMR
SR(INFMIN) = SORT(SECL)
IF (SR(INFMIN).LT.FDIFER) GO TO 87
INF = INFMIN
CALL FACTIR
IF (CRUCE.LT.1.0E-09) GO TO 80
F(INFMIN) = R(KC9)
GO TO 41
87 CONTINUE

```

```

C
C *****
C ** SE ENCUENTRA EL CENTROIDE DE LOS PUNTOS **
C ** PARA I DIFERENTE DE INFMAX **
C *****
C

```

```

DO 19 J=1,NVI
SUM2 = 0.
DO 20 I=1,N1
20 SUM2 = SUM2 + X2(I,J)
19 X2(N2,J) = 1./XN*(SUM2-X2(INFMAX,J))
SUM2 = 0.
DO 36 I=1,N1
DO 36 J=1,NVI
SUM2 = SUM2 + (X2(I,J) - X2(N2,J))**2
36 CONTINUE
FDIFER = (NVI + 1)/XN1*SORT(SUM2)
IF (FDIFER.LT.CRUCE) GO TO 98
FOIFER = CRUCE
GO TO 198
98 CRUCE = FOIFER
198 CONTINUE
FTER = F(INFMIN)
NCONT = NCONT + 1
IF (NCONT.LT.4*N1) GO TO 37
IF (ICONT.LT.1500) GO TO 337
CRUCF = 0.5*CRUCE
337 NCONT = 0.
WRITE(6,35)
WRITE(6,758) ICONT,FDIFER
CALL ESCRIB
IF (ICONT.GE.NUMITE) GO TO 9999
37 IF (FDIFER.LT.EPS) GO TO 81

```

```

C
C *****
C ** SELECCION DEL VALOR MAS PROXIMO AL MAXIMO VALOR **
C ** DE LA FUNCION OBJETIVO **
C *****
C

```

```

IF (INFMAX.EQ.1) GO TO 43
FS = F(1)
LSEG = 1
GO TO 44
43 FS = F(2)
LSEG = 2
44 DO 18 I=1,N1

```

```

IF(INFMAX.EQ.I) GO TO 18
IF(F(I).LT.FS) GO TO 18
FS = F(I)
LSEG = I
18 CONTINUE

```

```

*****
** REFLEXION DEL PUNTO QUE MAXIMIZA A TRAVES DEL CENTROIDE **
*****

```

```

DO 61 J=1,NVI
X2(N3,J) = X2(N2,J) + ALPHA*(X2(N2,J) - X2(INFMAX,J))
61 X(J) = X2(N3,J)
IN = N3
CALL SUMR
SR(N3) = SQRT(SECL)
IF(SR(N3).LT.FDIFER) GO TO 82
INF = N3
CALL FACTIB
IF(CRUCES.LT.1.0E-09) GO TO 80
82 CALL PRORLE(3)
F(N3) = R(KC9)
IF(F(N3).LT.F(INFMIN)) GO TO 84
IF(F(N3).LT.F(LSEG)) GO TO 92
GO TO 60
92 DO 93 J=1,NVI
93 X2(INFMAX,J) = X2(N3,J)
SR(INFMAX) = SR(N3)
F(INFMAX) = F(N3)
GO TO 1000

```

```

C
C *****
C ** VECTOR DE EUSQUEÑA A LO LARGO DE LA DIRECCION **
C ** A TRAVES DEL CENTROIDE Y EL VECTOR REFLEJADO **
C *****
C

```

```

84 DO 23 J=1,NVI
X2(N4,J) = X2(N3,J) + GAMMA*(X2(N3,J) - X2(N2,J))
23 X(J) = X2(N4,J)
IN = N4
CALL SUMR
SR(N4) = SQRT(SECL)
IF(SR(N4).LT.FDIFER) GO TO 25
INF = N4
CALL FACTIB
IF(CRUCES.LT.1.0E-09) GO TO 80
25 CALL PRORLE(3)
F(N4) = R(KC9)
IF(F(INFMIN).LT.F(N4)) GO TO 92
DO 26 J=1,NVI
26 X2(INFMAX,J) = X2(N4,J)
F(INFMAX) = F(N4)
SR(INFMAX) = SR(N4)
GO TO 1000
60 IF(F(N3).GT.F(INFMAX)) GO TO 64
DO 65 J=1,NVI
65 X2(INFMAX,J) = X2(N3,J)
64 DO 66 J=1,NVI
X2(N4,J) = BETA*X2(INFMAX,J) + (1. - BETA)* X2(N2,J)
66 X(J) = X2(N4,J)
IN = N4
CALL SUMR
SR(N4) = SQRT(SECL)
IF(SR(N4).LT.FDIFER) GO TO 67

```



```

764 FORMAT(//,10X,"EL VECTOR DETERMINADO POR PROGRAMA EL CUAL SATISFAC
4E",//,10X,"EL CRITERIO DE TOLERANCIA INICIAL ES=")
765 FORMAT(//,10X,"SUMA DE RESTRICCIONES VIOLADAS =",E17.7)
9999 STOP
END
SUBROUTINE FACTIB

```

```

C
C *****
C ** ESTA SUBROUTINA MINIMIZA LA SUMA DE LOS CUADRADOS DE LOS VALORES **
C ** DE LAS RESTRICCIONES VIOLADAS. ES LLAMADA CADA VEZ QUE EL VALOR **
C ** COMBINADO DE LAS RESTRICCIONES VIOLADAS EXCEDE EL CRITERIO DE **
C ** TOLERANCIA PARA LA ITERACION ACTUAL **
C *****
C

```

```

DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),
IRALD(100),R1(100),R3(100),FLG(10),H(50),SR(50)
COMMON/1/NVI,NRI,NRD,PASO,DUM1,DUM2,DUM3,IN,INF,FNIFER,SECL,KC1,
1KC2,KC3,KC4,KC5,KC6,KC7,KC8,KC9,X,X1,X2,R,SUM,F,SR,IRALD,ESCALA,
2CRUCE

```

```
COMMON/2/LFEAS,M5,M6,M7,M8,M9,S1A,S2A,S3A,TAM
```

```

ALPHA = 1.
BETA = 0.5
GAMMA = 2.
XNX = NVI
ICONT = 0
LCHEK = 0
ICHEK = 0

```

```

25 CALL COMENZ
DO 3 I=1,KC1
DO 4 J=1,NVI
4 X(J) = X1(I,J)
IN = I
CALL SUMR
3 CONTINUE

```

```

C
C *****
C ** SELECCIONA EL MAXIMO VALOR DE SUM(I) EN EL POLIEDRO
C *****
C

```

```

28 SUMH = SUM(1)
INDICE = 1
DO 7 I=2,KC1
IF(SUM(I).LE.SUMH) GO TO 7
SUMH = SUM(I)
INDICE = I
7 CONTINUE

```

```

C
C *****
C ** SELECCIONA EL MINIMO VALOR DE SUM(I) EN EL POLIEDRO
C *****
C

```

```

SUML = SUM(1)
KONT = 1
DO 8 I=2,KC1
IF(SUML.LE.SUM(I)) GO TO 8
SUML = SUM(I)
KONT = I
8 CONTINUE

```

```

C
C *****
C ** ENCUENTRA EL CENTROIDE DE LOS PUNTOS CON I DIFERENTE DE INDICE **
C *****
C

```

```
DO 9 J=1,NVI
```



```

SUM2 = 0.
DO 10 I=1,KC1
10 SUM2 = SUM2 + X1(I,J)
  X1(KC2,J) = 1./XNX*(SUM2 - X1(INDICE,J))
C
C *****
C ** ENCUENTRA LA REFLEXION DEL MAXIMO PUNTO A TRAVES DEL CENTROIDE **
C *****
C
  X1(KC3,J) = 2.*X1(KC2,J) - X1(INDICE,J)
9 X(J) = X1(KC3,J)
  IN = KC3
  CALL SUMR
  IF(SUM(KC3).LT.SUML) GO TO 11
C
C *****
C ** SELECCIONA EL VALOR MAS PROXIMO AL VALOR MAXIMO EN EL POLIEDRO
C *****
C
  IF(INDICE.EQ.1) GO TO 38
  SUMS = SUM(1)
  GO TO 39
38 SUMS = SUM(2)
39 DO 12 I=1,KC1
  IF((INDICE - 1).EQ.C) GO TO 12
  IF(SUM(I).LE.SUMS) GO TO 12
  SUMS = SUM(I)
12 CONTINUE
  IF(SUM(KC3).GT.SUMS) GO TO 13
  GO TO 14
C
C *****
C ** GENERA UN NUEVO MINIMO SI LA REFLEXION PRODUCE UN MINIMO **
C *****
C
11 DO 15 J=1,NVI
  X1(KC4,J) = X1(KC2,J) + 2.*(X1(KC3,J) - X1(KC2,J))
15 X(J) = X1(KC4,J)
  IN = KC4
  CALL SUMR
  IF(SUM(KC4).LT.SUML) GO TO 16
  GO TO 14
13 IF(SUM(KC3).GT.SUMH) GO TO 17
  DO 18 J=1,NVI
18 X1(INDICE,J) = X1(KC3,J)
17 DO 19 J=1,NVI
  X1(KC4,J) = 0.5*X1(INDICE,J) + 0.5*X1(KC2,J)
19 X(J) = X1(KC4,J)
  IN = KC4
  CALL SUMR
  IF(SUMH.GT.SUM(KC4)) GO TO 6
C
C *****
C ** REDUCE EL POLIEDRO POR MITAD SI LA REFLEXION PRODUCE UN VALOR **
C ** MAYOR QUE EL MAXIMO **
C *****
C
DO 20 J=1,NVI
DO 20 I=1,KC1
20 X1(I,J) = 0.5*(X1(I,J) + X1(KONT,J))
DO 29 I=1,KC1
DO 30 J=1,NVI
30 X(J) = X1(I,J)
  IN = I

```

```

CALL SUMR
29 CONTINUE
5  SUML = SUM(1)
   KONT = 1
   DO 23 I=2,KC1
   IF(SUML.LT.SUM(I)) GO TO 23
   SUML = SUM(I)
   KONT = I
23  CONTINUE
   SR(INF) = SORT(SUM(KONT))
   DO 27 J=1,NVI
27  X(J) = X1(KONT,J)
   GO TO 26
6   DO 31 J=1,NVI
31  X1(INDICE,J) = X1(KC4,J)
   SUM(INDICE) = SUM(KC4)
   GO TO 5
16  DO 21 J=1,NVI
   X1(INDICE,J) = X1(KC4,J)
21  X(J) = X1(INDICE,J)
   SUM(INDICE) = SUM(KC4)
   SR(INF) = SORT(SUM(KC4))
   GO TO 26
14  DO 22 J=1,NVI
   X1(INDICE,J) = X1(KC3,J)
22  X(J) = X1(INDICE,J)
   SUM(INDICE) = SUM(KC3)
   SR(INF) = SORT(SUM(KC3))
26  ICONT = ICONT + 1
   DO 36 J=1,NVI
36  XZ(INF,J) = X(J)
   IF(ICONT.LT.2*KC1) GO TO 50
   ICONT = 0
   DO 24 J=1,NVI
24  X(J) = X1(KC2,J)
   IN = KC2
   CALL SUMR
   DIFER = 0.
   DO 57 I=1,KC1
57  DIFER = DIFER + (SUM(I) - SUM(KC2))*2
   DIFER = 1./(KC7*NXN)*SORT(DIFER)
   IF(DIFER.GT.1.0E-14) GO TO 50

```

```

C
C *****
C ** SI EL POLIEDRO REGULABLE FALLA PARA SATISFACER LAS **
C ** RESTRICCIONES DENTRO DEL CRITERIO DE TOLERANCIA **
C ** EN LA ITERACION ACTUAL. LA BÚSQUEDA ES CAMBIADA DE LA POSICION **
C ** DONDE EL VECTOR X ES PUESTO Y ENTONCES FACTOR ES REPETIDA DESDE **
C ** EL PRINCIPIO *****
C

```

```

51 IN = KC1
   PASO = 20.*FDIFER
   CALL SUMR
   SR(INF) = SORT(SECL)
   DO 52 J=1,NVI
52  X1(KC1,J) = X(J)
   DO 53 J=1,NVI
   FACTOR = 1.
   X(J) = X1(KC1,J) + FACTOR*PASO
   X1(M9,J) = X(J)
   IN = M9
   CALL SUMR
   X(J) = X1(KC1,J) - FACTOR*PASO

```

```

X1(M5,J) = X(J)
IN = M5
CALL SUMR
56 IF(SUM(M9).LT.SUM(KC1)) GO TO 54
   IF(SUM(M5).LT.SUM(KC1)) GO TO 55
   GO TO 97
54 X1(M5,J) = X1(KC1,J)
   SUM(M5) = SUM(KC1)
   X1(KC1,J) = X1(M9,J)
   SUM(KC1) = SUM(M9)
   FACTOR = FACTOR + 1.
   X(J) = X1(KC1,J) + FACTOR*PASO
   IN = M9
   CALL SUMR
   GO TO 56
55 X1(M9,J) = X1(KC1,J)
   SUM(M9) = SUM(KC1)
   X1(KC1,J) = X1(M5,J)
   SUM(KC1) = SUM(M5)
   FACTOR = FACTOR + 1.
   X(J) = X1(KC1,J) - FACTOR*PASO
   IN = M5
   CALL SUMR
   GO TO 56

```

```

C
C *****
C ** SE EFECTUA UNA BUSQUEDA POR SECCION DORADA A LO LARGO DE CADA **
C ** COORDENADA *****
C

```

```

97 H(J) = X1(M9,J) - X1(M5,J)
   X1(M6,J) = X1(M5,J) + H(J)*S1A
   X(J) = X1(M6,J)
   IN = M6
   CALL SUMR
   X1(M7,J) = X1(M5,J) + H(J)*S2A
   X(J) = X1(M7,J)
   IN = M7
   CALL SUMR
   IF(SUM(M6).GT.SUM(M7)) GO TO 68
   X1(M8,J) = X1(M5,J) + (1. - S3A)*H(J)
   X1(M5,J) = X1(M7,J)
   X(J) = X1(M8,J)
   IN = M8
   CALL SUMR
   IF(SUM(M8).GT.SUM(M6)) GO TO 76
   X1(M5,J) = X1(M6,J)
   SUM(M5) = SUM(M6)
   GO TO 75
76 X1(M9,J) = X1(M8,J)
   SUM(M9) = SUM(M8)
   GO TO 75
68 X1(M9,J) = X1(M6,J)
   X1(M8,J) = X1(M5,J) + S3A*H(J)
   X(J) = X1(M8,J)
   IN = M8
   CALL SUMR
   PASO = TAM
   SUM(M9) = SUM(M6)
   IF(SUM(M7).GT.SUM(M8)) GO TO 71
   X1(M5,J) = X1(M8,J)
   SUM(M5) = SUM(M8)
   GO TO 75
71 X1(M9,J) = X1(M7,J)

```

```

SUM(M9) = SUM(M7)
75 IF (ABS(X1(M9,J) - X1(M5,J)).GT.0.01*FDIFER) GO TO 97
X1(KC1,J) = X1(M7,J)
X(J) = X1(M7,J)
SUM(KC1) = SUM(M5)
SR(INF) = SORT(SUM(KC1))
IF (SR(INF).LT.FDIFER) GO TO 760
53 CONTINUE
ICHEK = ICHEK + 1
PASO = FDIFER
IF (ICHEK.LE.2) GO TO 25
CRUCE = 1.0E-12
WRITE(6,853)
WRITE(6,850)
WRITE(6,851) (X(J),J=1,NVI)
WRITE(6,852) FDIFER,SR(INF)
GO TO 46
760 DO 761 J=1,NVI
X2(INF,J) = X1(KC1,J)
761 X(J) = X1(KC1,J)
50 IF (SR(INF).GT.FDIFER) GO TO 28

```

```

C
C *****
C ** MODIFICA LA INTERPOLACION DE LAGRANGE PARA DESIGUALDADES **
C ** AJUSTADAS
C *****
C

```

```

IF (SR(INF).GT.0.) GO TO 35
CALL PROBLE(3)
FINT = R(KC9)
DO 139 J=1,NVI
139 X(J) = X2(INF,J)
CALL PROBLE(2)
DO 40 J=KC7,KC8
40 R1(J) = R(J)
DO 41 J=1,NVI
41 X(J) = X1(KONT,J)
CALL PROBLE(2)
DO 42 J=KC7,KC8
42 R3(J) = R(J)
DO 43 J=1,NVI
H(J) = X1(KONT,J) - X2(INF,J)
43 X(J) = X2(INF,J) + 0.5*H(J)
CALL PROBLE(2)
FLG(1) = 0.
FLG(2) = 0.
FLG(3) = 0.
DO 44 J=KC7,KC8
IF (R3(J).GE.0.) GO TO 44
FLG(1) = FLG(1) + R1(J)**2
FLG(2) = FLG(2) + R(J)**2
FLG(3) = FLG(3) + R3(J)**2
44 CONTINUE
SR(INF) = SORT(FLG(1))
IF (SR(INF).LT.FDIFER) GO TO 35
ALPHA1 = FLG(1) - 2.*FLG(2) + FLG(3)
BETA1 = 3.*FLG(1) - 4.*FLG(2) + FLG(3)
RADIO = BETA1/(4.*ALPHA1)
DO 45 J=1,NVI
45 X(J) = X2(INF,J) + H(J)*RADIO
IN = INF
CALL SUMR
SR(INF) = SORT(SECL)
IF (SR(INF).LT.FDIFER) GO TO 465

```



```

2 FORMAT (//,10X,"LOS VALORES DE LOS VECTORES INDEPENDIENTES SON",//,
110X,6E17.7)
IF (NR1.EQ.0) GO TO 6
CALL PROBLE(1)
WRITE(6,3) (R(J),J=1,NR1)
3 FORMAT (//,10X,"LOS VALORES DE LAS RESTRICCIONES QUE SON IGUALDADES
1 SON",//,10X,6E17.7)
6 IF (NRD.EQ.0) GO TO 5
CALL PROBLE(2)
WRITE(6,4) (R(J),J=KC7,KC6)
4 FORMAT (//,10X,"LOS VALORES DE LAS RESTRICCIONES QUE SON DESIGUALDA
IDES SON",//,6E17.7)
5 RETURN
END
SUBROUTINE SUMR

```

```

C
C *****
C ** ESTA SUBROUTINA CALCULA LA SUMA DE CUADRADOS DE LOS VALORES **
C ** DE LAS RESTRICCIONES PARA SER COMPARADA CON EL CRITERIO DE **
C ** TOLERANCIA **
C *****
C
C DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),
1SR(50),RALD(100)
COMMON/1/NVI,NR1,NRD,PASO,ALPHA,BETA,GAMMA,IN,INF,FDIFER,SECL,KC1,
2KC2,KC3,KC4,KC5,KC6,KC7,KC8,KC9,X,X1,X2,R,SUM,F,SR,RALD,
3CRUCE
COMMON/2/LFEAS,M5,M6,M7,M8,M9,S1A,S2A,S3A,TAN
SUM(IN) = 0.
CALL PROBLE(2)
SECL = 0.
IF (NRD.EQ.0) GO TO 4
DO 1 J=KC7,KC8
IF (R(J).GE.0.) GO TO 1
SECL = SECL + R(J)**2
1 CONTINUE
4 IF (NR1.EQ.0) GO TO 3
CALL PROBLE(1)
DO 2 J=1,NR1
2 SECL = SECL + R(J)**2
3 SUM(IN) = SECL
5 RETURN
END
SUBROUTINE PROBLE(INQ)
DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),
1SR(50),RALD(100)
COMMON/1/NVI,NR1,NRD,PASO,ALPHA,BETA,GAMMA,IN,INF,FDIFER,SECL,KC1,
2KC2,KC3,KC4,KC5,KC6,KC7,KC8,KC9,X,X1,X2,R,SUM,F,SR,RALD,
3CRUCE
COMMON/2/LFEAS,M5,M6,M7,M8,M9,S1A,S2A,S3A,TAN
IF (INQ.EQ.3) GO TO 3
IF (INQ.EQ.2) GO TO 2
C
C *****
C ** RESTRICCIONES QUE SON IGUALDADES **
C *****
C
C R(1) = X(1)**2 + X(2)**2 - 9.*X(2) + 4.25
GO TO 5
C
C *****
C ** RESTRICCIONES QUE SON DESIGUALDADES **
C *****
C

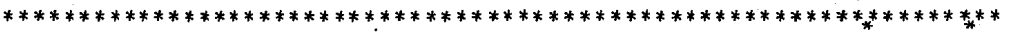
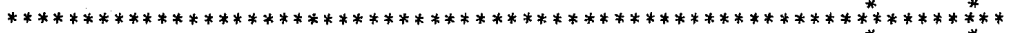
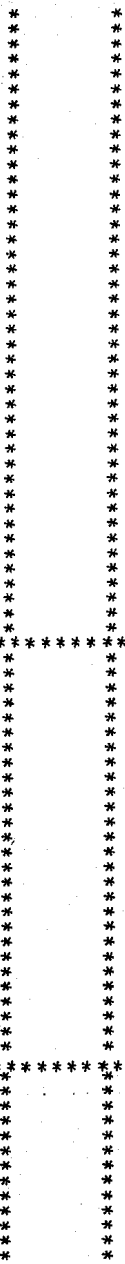
```

```
2 R(2) = X(1)
R(3) = X(2)
GO TO 5
```

```
C
C *****
C ** FUNCION OBJETIVO **
C *****
C
```

```
3 R(4) = X(1)**2 + X(2)**2
5 RETURN
END
```

BIBLIOGRAFIA



B I B L I O G R A F I A

1. Avriel, M., "GENERALIZED CONVEX FUNCTIONS WITH APPLICATIONS TO NONLINEAR PROGRAMMING", Mathematical Programs for Activity Analysis, 1974.
2. Bector, C. R., "ON CONVEXITY, PSEDO-CONVEXITY AND QUASI-CONVEXITY", Cahiers Centre Etudes Recherche Operationnelle, 1973.
3. Bhatt, S. K., "SUFFICIENT OPTIMALITY CRITERIA IN NONLINEAR PROGRAMMING IN PRESENCE OF CONVEX EQUALITY AND INEQUALITY CONSTRAINTS", Zeitschrift für Operations Research, 19, pags. 101-105, 1975.
4. Dubois, J., "THEOREMS OF CONVERGENCE FOR IMPROVED NONLINEAR PROGRAMMING ALGORITHMS", Operations Research, 21, pags. 328-332, 1973.
5. Fiacco, A. V., "CONVERGENCE PROPERTIES OF LOCAL SOLUTIONS OF SEQUENCES OF MATHEMATICAL PROGRAMMING PROBLEMS IN GENERAL SPACES", J. Optimization Theory and Applications, 13, pags.1-12, 1974.
6. Gehner, K. R., "NECESSARY AND SUFFICIENT OPTIMALITY CONDITIONS FOR FRITZ JOHN PROBLEM WITH LINEAR EQUALITY CONSTRAINTS", SIAM J. Control, 12, pags. 140-149, 1974.
7. Greenberg, H.J., "A REVIEW OF QUASI-CONVEX FUNCTIONS", Operations Research, 29, pags. 1553-1570, 1971.
8. Hogan, W. W., "POINT-TO-SET MAPS IN MATHEMATICAL PROGRAMMING", SIAM Reviw, 15, pags. 591-603, 1973.

9. Huard, P., "OPTIMIZATION ALGORITHMS AND POINT-TO-SET MAPS", *Mathematical Programming*, 8, pags. 308-331, 1975.
10. Meyer, R. R., "SUFFICIENT CONDITIONS FOR THE CONVERGENCE OF MONOTONIC MATHEMATICAL PROGRAMMING ALGORITHMS", *J. Computer and Systems Sciences*, 12, pags 108-121, 1976.
11. Paviani, T. D., "CONSTRAINED NONLINEAR OPTIMIZATION", *Operations Research*, 1980.
13. Polak, E., "ON THE IMPLEMENTATION OF CONCEPTUAL ALGORITHMS", Ritter (Eds.), 1970.
14. Powell, M. J. D., "SOME GLOBAL CONVERGENCE PROPERTIES OF A VARIABLE METRIC ALGORITHM FOR MINIMIZATION", American Mathematical Society, Providence, R. I., 1976.
15. Robinson, S. M., "A SUFFICIENT CONDITION FOR CONTINUITY OF OPTIMAL SETS IN MATHEMATICAL PROGRAMMING". *J. Mathematical Analysis and Applications*, 45, pags. 506-511, 1974.
17. Rockafeller, R. T., " CONVEX ANALYSIS", Princeton University Press, Princeton, N. J., 1970.
18. Schaible, S., "QUASI-CONCAVE, STRICTLY QUASI-CONCAVE AND PSEUDO-CONCAVE FUNCTIONS", *Operations Research Verfahren*, 17, pags 308-316, 1973.
19. Wolfe, P., "CONVERGENCE THEORY IN NONLINEAR PROGRAMMING", J. Abadie (Eds.), 1970.
20. Zangwill, W. I., "NONLINEAR PROGRAMMING: A UNIFIED APPROACH", Prentice-Hall, Englewood Cliffs, N.J., 1969.