UNIVERSIDAD DE EL SALVADOR

Facultad de Ingeniería y Arquitectura
Departamento de Matemática

INTRODUCCION A LA TEORIA DE APROXIMACION

TRABAJO DE GRADUACION

PRESENTADO POR

CRISTINA CUBIAS LOPEZ

PREVIO A LA OPCION DEL TITULO DE

Licenciada en Matemática

- JUNIO 1979 -

SAN SALVADOR, EL SALVADOR, CENTRO AMERICA

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
DEPARTAMENTO DE MATEMATICA

TRABAJO DE GRADUACION

"INTRODUCCION A LA TEORIA DE APROXIMACION"

JUNIO DE 1979

SAN SALVADOR, EL SALVADOR, CENTRO AMÉRICA

UNIVERSIDAD DE EL SALVADOR

RECTOR:

DR. EDUARDO BADÍA SERRA

SECRETARIO GENERAL:

DR. JORGE FERRER DENIS

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO:

ING. EDUARDO CASTILLO URRUTIA

SECRETARIO:

ING. JUAN MIGUEL IGLESIAS CARRANZA

DEPARTAMENTO DE MATEMATICA

JEFE DEL DEPARTAMENTO:

ING. GABRIEL MELENDEZ MAYORGA

TRABAJO DE GRADUACION DESARROLLADO POR:

CRISTINA CUBÍAS LÓPEZ

PREVIO A LA OPCIÓN DE SU TITULO DE:

LICENCIADA EN MATEMÁTICA

TRABAJO DE GRADUACION

ASESORES:

LIC. MAURO HERNÁN HENRÍQUEZ RAUDA

ING. GABRIEL MELÉNDEZ MAYORGA

A mis padres y a mi hermana,
con agradecimiento;
A mi hermano,
con cariño.

INTRODUCCION

La idea de investigar sobre la TEORIA DE APROXIMACION, y, lo más importante, el deseo de hacerlo, surgió de la motivación producida por una plática sostenida con un grupo de compañeros, en la cual comentábamos que este tópico está poco explorado hasta hoy, en nuestro medio y en nuestra carrera; no obstante su gran importancia en la resolución de ciertos problemas prácticos de la computación.

Es apropiado considerar algunos ejemplos concretos de tales problemas. En estos ejemplos se observará una similitud básica en que cada uno involucra la selección de una clase prescrita de funciones de un elemento que es, en algún sentido, próximo a una cierta función fija.

1) Determinar un polinomio p de grado mínimo tal que, sobre el intervalo $[0;\pi/2]$ tendremos $|p(x) - \text{sen } x| \leq 10^{-8}$.

2) Más general, dada una función f y un número positivo ε, determinar un polinomio p tal que $|p(x) - f(x)| \leq \varepsilon$ en algún intervalo $[a,b]$

3) Determinar un vector $x = (x_1, \ldots, x_n)$ que sea la mejor solución aproximada (en el sentido de los mínimos cuadrados), al sistema de ecuaciones lineales:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i (i = 1, \ldots, m).$$

Esto es, determinar x que minimize la expresión:

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2$$

4) Como una variante de (3), podemos pedirle a x que minimize:

$$\max_{1 \leq i \leq m} \left| \Sigma a_{ij} x_j - b_j \right|$$

5) como variante de (2) podemos requerir que

$$\int_a^b |p(x) - f(x)|^2 \, dx < \varepsilon$$

Estos son problemas típicos de computación en la TEORIA DE APROXIMACION. Gran cantidad de preguntas de interés matemático más profundo surgen en una forma natural a partir de los problemas citados. Por ejemplo, podemos preguntar: ¿Puede un polinomio ser encontrado siempre en la solución de un problema (1)?, ¿Cuál es exactamente la clase de funciones f - para la cual el problema (2) puede ser siempre resuelto (esto es para todo $\varepsilon > 0$)?. Para una función fija en el problema (2), ¿Cómo se comporta el grado del polinomio cuando el ε tiende a cero?, ¿Qué relación existe- entre los polinomios que resuelven (2) y (5). Si el problema (2) se vuelve extremadamente laborioso, ¿Existen otras aproximaciones que permitan- resolverlo más facilmente aunque no sean muy óptimas?

El presente trabajo de investigación contesta adecuadamente estas - cuestiones teóricas a lo largo de su desarrollo, con lo cual ha sido satisfecha mi motivación, y pienso además, que al menos en la forma más - sencilla, sirva de incentivo para que otros más experimentados dirijan - su atención a ésta área del estudio de la matemática en beneficio del desarrollo de nuestro departamento.

INDICE

PAG.

PAG.

CAPITULO 1

LAS SOLUCIONES TCHEBYCHEFF DE ECUACIONES LINEALES INCONSISTENTES

1. INTRODUCCION

En este capítulo consideraremos algunos problemas de aproximación, la cual proviene de un sistema de ecuaciones lineales.

$$(1) \qquad \sum_{j=1}^{n} A_j^i x_j = b_i \qquad (i = 1,\ldots m)$$

Vamos a suponer que los datos A_j^i y b_i son conocidos, y que las incognitas x_j deben der determinadas. Así, un sistema tal como (1) nos presenta un problema de aproximación: para determinar sus soluciones aproximadas o exactas. Un sistema de ecuaciones lineales pueda que no tenga solución, exactamente una solución o infinitamente muchas soluciones, dependiendo de los datos. En primer lugar, uno podría creer que el caso, cuando no existe solución, es el menos interesante de los tres y exceto de significado práctico. Pero resulta que lo opuesto es verdadero: El cálculo de una solución aproximada para (1) cuando no existe solución exacta es un problema no trivial e importante. Muchos problemas prácticos de aproximación se reducen a uno de este tipo, o algunas veces a una sucesión de tales problemas.

Empezaremos con un teorema que indica como los datos de (1) de-

terminan la naturaleza de las soluciones. Para cada $j = 1, \ldots, n$ - sea A_j el vector $[A_j^1, A_j^2, \ldots, A_j^m]$. También sea b el vector

$$[b_1, b_2, \ldots, b_m].$$

El sistema (1) puede ahora ser escrito en la forma

$$\sum_{j=1}^{n} x_j A_j = b.$$

Si esta ecuación es consistente (es decir, posee una solución), entonces b está situada en el espacio lineal generado por los vectores $A_1, \ldots, A_n$.

TEOREMA: El sistema de ecuaciones lineales $\sum_{j=1}^{n} A_j^i x_j = b_i (i=1, \ldots, m)$ es consistente si y sólo si b está situado en el espacio lineal generado por los vectores A_j. Si el sistema es consistente, tiene exactamente una solución si y sólo si los A_j son linealmente independientes.

Si el sistema (1) es inconsistente (es decir, que no posee solución exacta), entonces podemos, sin embargo, buscar minimizar las discrepancias entre los números b_i y los números $\Sigma A_j^i x_j$. Viendo esto - desde otro punto de vista, podemos preguntar si el vector $\Sigma x_j A_j - b$ (el cual no puede ser cero), estará en algún sentido cerca del cero. Dicho todavía en otra forma, el problema es determinar las x_j tal - que el vector $\Sigma x_j A_j$ esté tan cerca como sea posible de b.

Del teorema de la existencia general, (ver Apéndice T.1) resulta que para cualquier norma definida en E_m existe una solución para-

este problema. En otras palabras, existe un vector $x = [x_1, \ldots, x_n]$ para el cual la expresión $||\Sigma x_j A_j - b||$ es un mínimo. Tal vector puede ser llamado una solución mejor aproximada del sistema (1). Generalmente habrá diferentes soluciones mejor aproximadas para diferentes selecciones de la norma, y aun para una norma particular puede que haya un conjunto grande de soluciones mejor aproximadas. En este capítulo ponemos especial énfasis en el problema asociado con la norma:

$$||y||_T = \max_{1 \le i \le m} |y_i|$$

Cuando esta norma es empleada, una solución mejor aproximada de (1) da a la expresión

$$\Delta(x) = \max_{1 \le i \le m} \left| \sum_{j=1}^{n} A_j^i x_j - b_i \right| \tag{2}$$

un mínimo. Tal x es por lo tanto algunas veces llamada una solución minimax. Ya que P. L Tchebycheff fue el primero en realizar una investigación sistemática de las aproximaciones minimax, x es también llamada una aproximación Tchebycheff y la norma, la norma Tchebycheff. Esto explica el suscrito T.

La presente discusión será simplificada algunas veces considerando un problema ligeramente diferente, el cual abarca el problema Tchebycheff anteriormente descrito.

En lugar de minimizar la expresión (2), minimizamos la expresión

$$\delta(x) = \max_{i} \left(\sum_{j=1}^{n} A_j^i x_j - b_i \right) \tag{3}$$

La manera en la cual el problema Tchebycheff es incluído por (3) es como sigue: Doblamos el número de datos colocando

$$A^{i+m} = - A_i \quad y \quad b_{i+m} = - b_i$$

Así, si ponemos

$$r_i = \sum_{j=1}^{n} A_j^i x_j - b_i, \text{ entonces } r_{i+m} = - r_i$$

y,

$$\max_{1 \leq i \leq m} |r_i| = \max_{1 \leq i \leq 2m} r_i.$$

Por supuesto, la función $\delta(x)$ definida en (3) no podría ser acotada inferiormente. Pero si esta función surge de un problema Tchebycheff en la forma justamente indicada, entonces será acotada inferiormente y obtendrá su mínimo. Será conveniente considerar las funciones de la forma δ las cuales son inferiormente acotadas por cero.

2. SISTEMAS DE ECUACIONES CON UNA INCOGNITA

En esta sección discutimos varios procedimientos para resolver el problema Tchebycheff conectados con un sistema de ecuaciones:

$$a_i x = b_i (i = 1,\ldots,m),$$

siendo este el caso $n = 1$ del sistema (1) considerado anteriormente. La razón para emplear este caso simple es que nuestros algoritmos pueden ser usados como componentes de algoritmos más sofisticados. Además la familiaridad con el caso más simple hará más transparente el caso general.

Comenzamos con un ejemplo idealizado pero concreto, el cual ilustra cómo un problema Tchebycheff puede surgir en la práctica.

Suponga que se desea estimar la asi llamada "Constante Móvil" de un móvil, midiendo la prolongación producida por varias fuerzas. Digamos que los resultados de la experimentación son como sigue

x (fuerza)	2.0	4.0	5.0	6.0
y (prolongación)	1.2	2.1	2.6	3.1

Por la ley de Hook, $y = cx$. Asi, cada observación es capaz de producir un valor de c, pero a causa de los errores en la medición, estos valores de c no están de acuerdo. Tenemos en realidad el siguiente sistema inconsistente de ecuaciones lineales para la determinación de c:

$$2.0\ c = 1.2$$
$$4.0\ c = 2.1$$
$$5.0\ c = 2.6$$
$$6.0\ c = 3.1$$

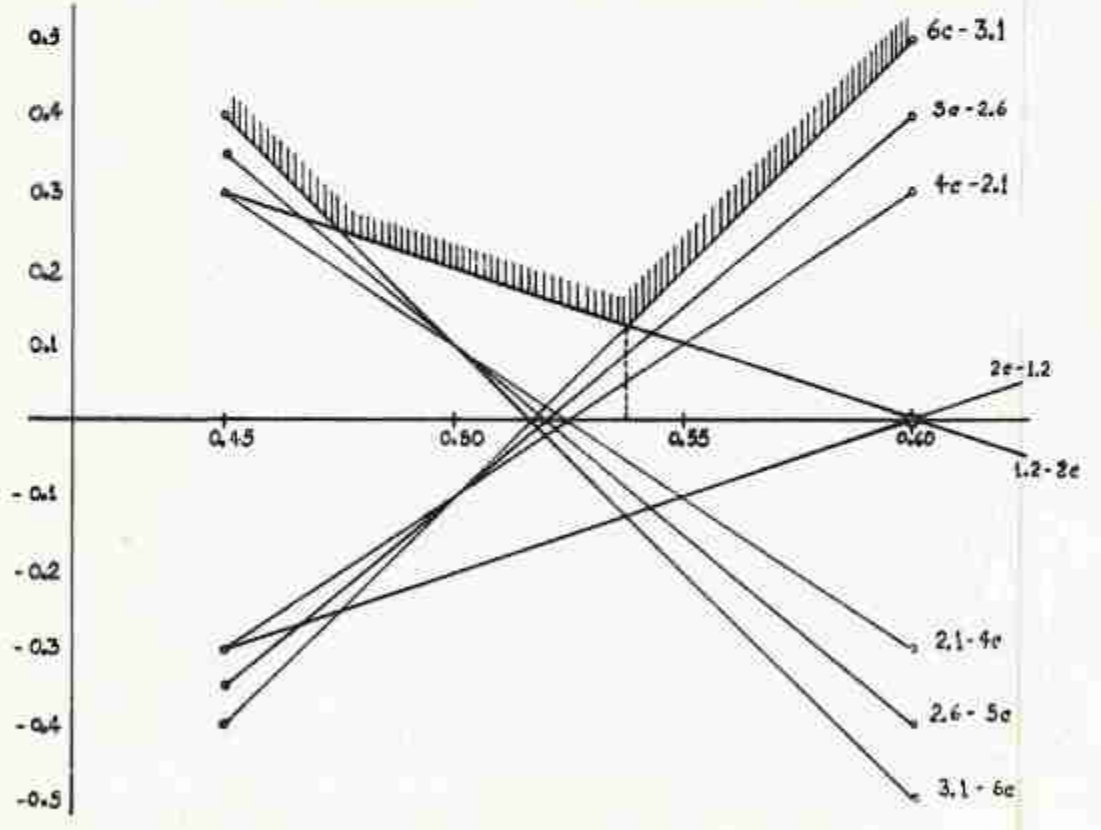
Estamos de acuerdo entonces que c será seleccionado como una so-

lución Tchebycheff de este sistema. Entonces buscamos minimizar la expresión.

$$\Delta(c) = \max\{|2c - 1.2|, |4c - 2.1|, |5c - 2.6|, |6c - 3.1|\}$$

Comenzamos graficando las ocho líneas rectas, $v = 2c - 1.2$, $v = -2c + 1.2$, $v = 4c - 2.1$, $v = -4c + 2.1$, etc. Despues que esto ha sido hecho, la línea quebrada que alcanza la altura máxima es el gráfico de $\Delta(c)$.

El resultado se muestra en la figura.



Del bosquejo podemos estimar que el mínimo ocurre para $c = 0.54$ y es aproximadamente 0.11. En efecto, el mínimo ocurre en la intersección de dos rectas y es obtenible exactamente resolviendo la ecuación

$$6c - 3.1 = 1.2 - 2c$$

De esta forma el valor correcto de c es 0.5375 y el valor correspondiente de Δ es 0.125.

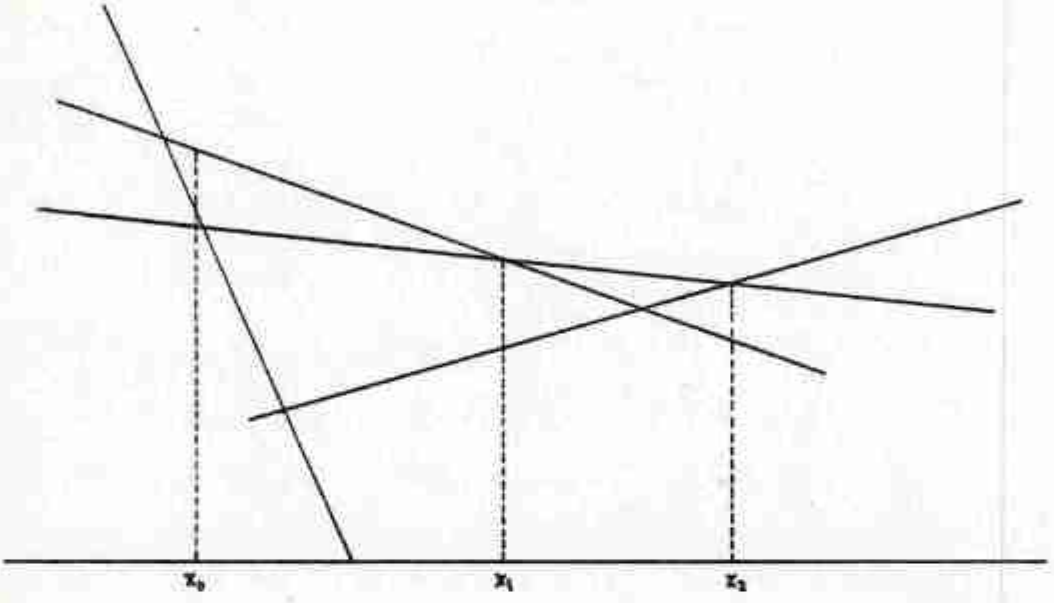
Antes de dejar este ejemplo observamos el gráfico de Δ su no diferenciabilidad y su convexidad. Una prueba de la convexidad de tal función se da mostrando que

$\Delta(\lambda c + \mu d) \leq \lambda\Delta(c) + \mu\Delta(d)$. También es notable que en la solución $c = 0.5375$, dos de las ecuaciones tenian errores que - eren iguales en magnitud (es decir 0.125) mientras que las ecuaciones - que quedaron tenian errores menores.

Una reflexión de momento nos muestra que esto se esperaría en cualquier problema similar, sin importar cuantas ecuaciones hubieran. En el problema general, con n incognitas, esperaremos $n + 1$ errores iguales en magnitud en la solución.

Ahora volvamos al problema de la construcción de los algoritmos para estos problemas. El método gráfico sugerido mediante el ejemplo anterior, es un algoritmo que es seguro; pero no es capaz de ser automatizado. Además, una extensión para n variables parece ser bastante dificil. Como fue señalado en la sección 1, es un problema más general buscar el punto mínimo para una función de la forma

$$\delta(x) = \max_{1 \leq i \leq m} \{a_i x - b_i\}$$



Ejemplo:Encuentre la solución Tchebycheff para el siguiente sistema:

$$r_1(x) = -3x + 6,$$
$$r_2(x) = -\frac{3}{11}x + 3,$$
$$r_3(x) = -0.79x + 4.6,$$
$$r_4(x) = \frac{x}{2} - \frac{1}{4}$$

SOLUCION: Ver gráfico siguiente.

Sea $x_o = 0.25$

Definamos $M = \{i/r_i(0.25) = \delta(0.25)\}$. Entonces $M = \{1\}$

Como $a1 = -3 < 0$, disminuimos $r_1(x)$ hacia la derecha hasta en-

contrar un vértice

$$r_1(x_1) = r_3(x_1)$$

$$-3x + 6 = -0.79x + 4.6$$

entonces $x = 0.63$.

Sea $x_1 = 0.63$.

$$M = \{i/r_i(0.63) = \delta(0.63)\} = \{1,3\}$$

Como $a_1 = -3 < 0$ y $a_3 = -0.79$, $a_1a_3 < 0$. Por ser

$$\min \{|a_1|, |a_3|\} = |a_3|, \text{ seleccionamos } j = 3.$$

Como $a_3 < 0$, disminuimos $r_3(x)$ hacia la derecha hasta encontrar un vértice

$$r_3(x_2) = r_2(x_2)$$

$$-0.79x + 4.6 = -\frac{3}{11}x + 3,$$

entonces $x = 3.09$.

Sea $x_2 = 3.09$

$$M = \{i/r_i(3.09) = \delta(3.09)\} = \{2,3\}$$

Como $a_2 = -\frac{3}{11} < 0$ y $a_3 = -0.79 < 0$, $a_2a_3 > 0$. Por ser

$$\min \{|a_2|, |a_3|\} = |a_2|, \text{ seleccionamos } j = 2$$

Como $a_2 < 0$, disminuimos $r_2(x)$ hacia la derecha hasta encontrar un vértice

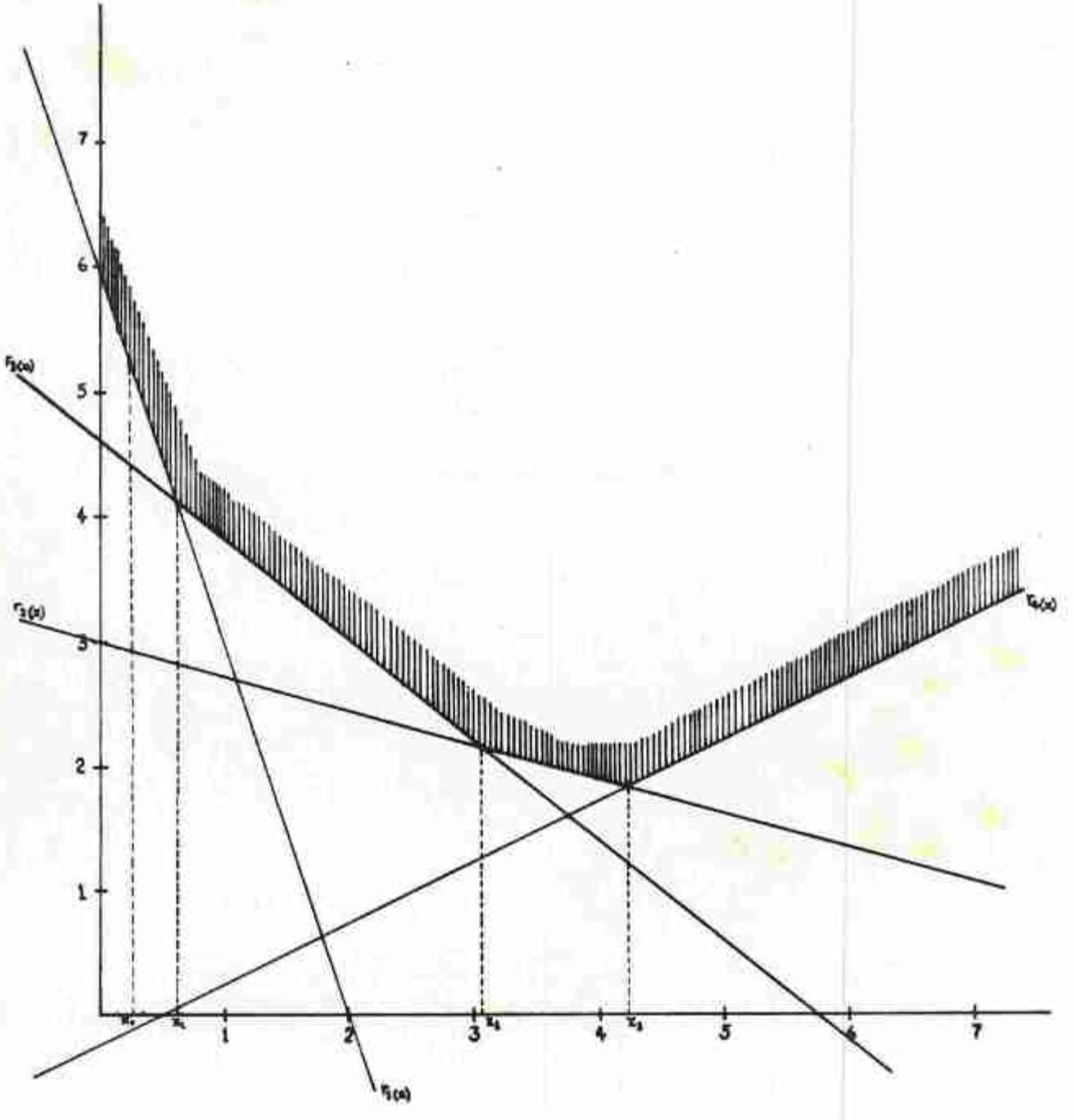
$$r_2(x_3) = r_4(x_3)$$

$$-\frac{3}{11}x + 3 = \frac{x}{2} - \frac{1}{4}, \text{ entonces } x = 4.2$$

Sea $x_3 = 4.2$

$$M = \{i/r_i(4.2) = \delta(4.2)\} = \{2,4\}$$

Dado que $a_2 = -\frac{3}{11} < 0$ y $a_4 = \frac{1}{2} > 0$, $a_2a_4 < 0$, lo cual implica que $x = 4.2$ es la solución.



$r_1(x) = -3x + 6$

$r_2(x) = -\frac{3}{11}x + 3$

$r3(x) = -0.79x + 4.6$

$r_4(x) = \frac{1}{2}x - \frac{1}{4}$

ALGORITMO 2. (Ascenso de vértice a vértice).

En cada paso de este algoritmo tenemos un punto x_o y un par de índices j y k tal que

$$r_j(x_o) = r_k(x_o),\ a_j \leq 0 \leq a_k \quad y \quad a_j \neq a_k.$$

Es fácil ver que bajo estas circunstancias x_o es un punto mínimo de la función

$$\max \{r_j(x), r_k(x)\}.$$

La gráfica de δ sugiere que por lo menos uno de los puntos mínimos de δ tendrá la misma propiedad para un par apropiado de índices. Ahora seleccione i tal que

$$r_i(\bar{x}) = \delta(x_o).$$

Si $a_i < 0$, avanzamos a la intersección de r_i y r_k.

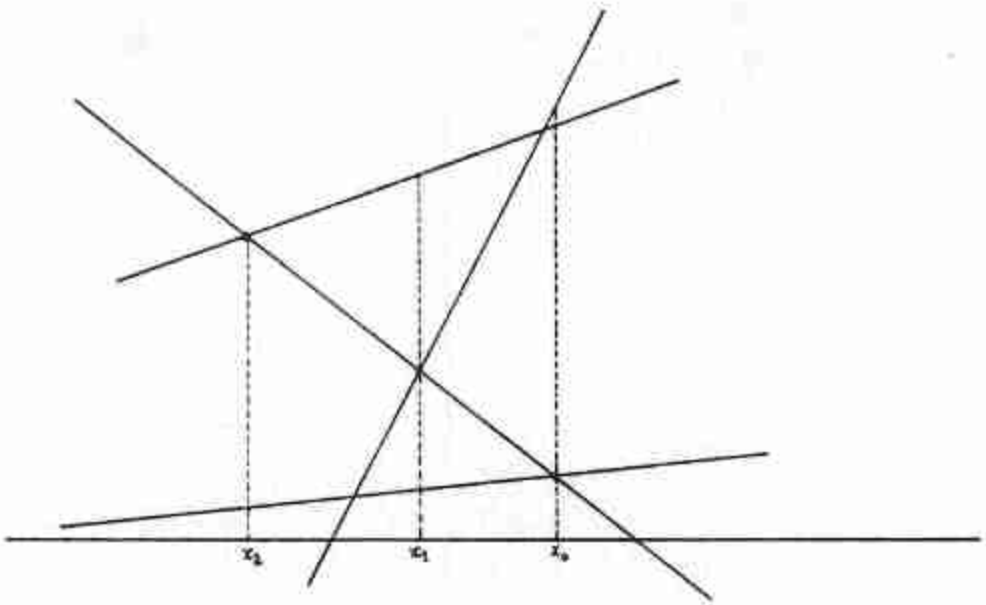
Si $a_i > 0$, avanzamos a la intersección de r_i y r_j.

Si $a_i = 0$, entonces avanzamos a la intersección de r_i con cualquiera de los r_j ó r_k que tenga un coeficiente no cero de x.

Cuando $a_i < 0$ el nuevo punto que sustituye a x_o tiene la forma

$$x = \frac{(b_k - b_i)}{(a_k - a_i)}$$

También sustituimos j por i y comenzamos de nuevo.



Ejemplo: Encuentre la solución Tchebycheff para el siguiente sistema:

$$r1(x) = \frac{3}{2}$$

$$r2(x) = \frac{1}{3}x$$

$$r3(x) = -\frac{12}{7}x + 12$$

$$r4(x) = -\frac{11}{9}x + 11$$

$$r5(x) = -\frac{9}{12}x + 9$$

SOLUCION: Sea $x_o = \frac{49}{8}$ y el par de índices 1 y 3 donde

$$r_1(\tfrac{49}{8}) = r_3(\tfrac{49}{8})$$

$a_1 = 0 \geq 0$, $a_3 = -\frac{12}{7} < 0$, $a_3 \leq 0 \leq a_1$, $a_1 \neq a_3$.

Es claro que x_o es punto mínimo de $\max\{r1(x), r_3(x)\}$

Seleccionamos el índice 5 porque $r_5(4\frac{9}{8}) = \delta(4\frac{9}{8})$.

Como $a_5 = -\frac{9}{12} < 0$, avanzamos a la intersección de

r_5 y r_1: $r_5(x_1) = r_1(x_1)$.

$-\frac{9}{12}x + 9 = \frac{3}{2}x$, entonces $x = 10$.

Sea $x_1 = 10$ y el par de índices 1 y 5, donde

$r_1(10) = r_5(10)$.

$a_1 = 0 \geq 0$; $a_5 = -\frac{9}{12} < 0$, $a_5 \leq 0 \leq a_1$, $a_5 \neq a_1$.

Seleccionamos el índice 2 porque $r_2(10) = \delta(10)$

Como $a_2 = \frac{1}{3} > 0$, avanzamos a la intersección de

r_2 y r_5: $r_2(x_2) = r_5(x_2)$

$\frac{1}{3}x = -\frac{9}{12}x + 9$, entonces $x = 8.3$

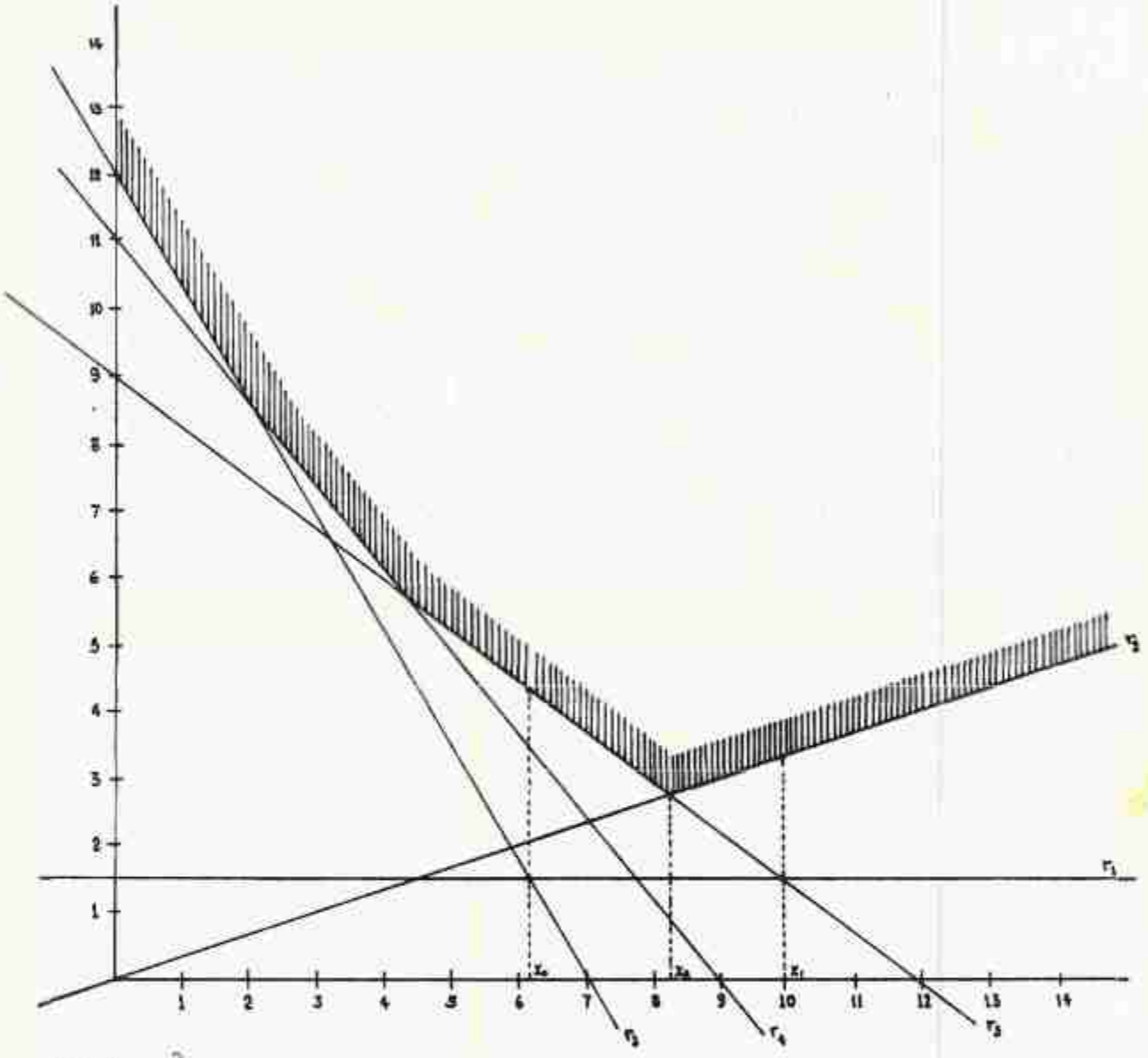
Sea $x_2 = 8.3$ y el par de índices 2 y 5 donde

$r_2(8.3) = r_5(8.3)$.

Como ya no es posible encontrar otro índice i tal que

$r_i(x_2) = \delta(x_2)$, x_2

es la solución.



$r_1(x) = \frac{2}{3}$

$r_2(x) = \frac{1}{3} x$

$r3(x) = -\frac{12}{7} x + 12$

$r_4(x) = -\frac{11}{5} x + 11$

$r_5(x) = -\frac{9}{12} x + 9$

ALGORITMO 3. (Investigación)

En cada paso de este algoritmo tenemos dos puntos x Λ y, estando situados en lados opuestos del punto mínimo. Sea $x < y$. Este estado de cosas será evidenciado como sigue:

Sea

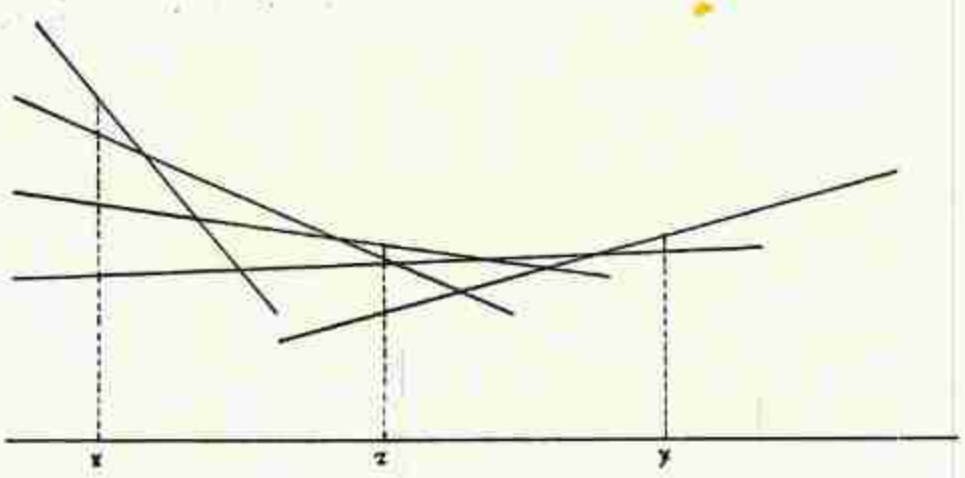
$$r_i(x) = \delta(x) \text{ y } r_j(y) = \delta(y).$$

entonces $a_i \leq 0 \leq a_j$.

Si $a_i = 0$, entonces x es una solución y si $a_j = 0$, y es una solución Ponga $z = 1/2(x+y)$. Suponga que

$$r_k(z) = \delta(z).$$

Si $a_k < 0$ sustituya x por z e i por K. Si $a_K > 0$, sustituya y por z y j por K. Comience de nuevo. En n pasos la exactitud con la cual el punto mínimo es localizado será mejorado mediante un factor de 2^{-n}.



Ejemplo: Investigue la localización del punto mínimo del siguiente sistema:

$$r_1(x) = \frac{2}{3}x - 4$$

$$r_2(x) = \frac{1}{5}x + 1$$

$$r_3(x) = -\frac{2}{9}x + 4$$

$$r_4(x) = -\frac{3}{4}x + 6$$

$$r_5(x) = -\frac{8}{5}x + 8$$

Sean $x = 1$ y $y = 15$ dos puntos situados en lados opuestos del punto mínimo z.

$1 < 15$. Con estas condiciones, tenemos:

$$r_5(1) = \delta(1) \quad y \quad r_1(15) = \delta(15)$$

Entonces

$$a_5 = -\frac{8}{5} \leq 0 \qquad a_1 = \frac{2}{3} \geq 0$$

Como $a_5 \neq 0$ entonces 1 no es solución.

Como $a_1 \neq 0$ entonces 15 no es solución.

Pongamos

$$z = \frac{1}{2}(1 + 15) = \frac{16}{2} = 8$$

$$r_2(8) = \delta(8)$$

Como

$$a_2 = \frac{1}{5} > 0$$

sustituimos 15 por 8 y el índice 1 por 2

Así

$$r_5(1) = \delta(1) \quad y \quad r_2(8) = \delta(8)$$

Entonces

$$a_5 = -\frac{8}{5} \leq 0 \qquad a_1 = \frac{1}{5} \geq 0$$

Como $a_5 \neq 0$ entonces 1 no es solución

Como $a_2 \neq 0$ entonces 8 no es solución.

Pongamos

$$m = \frac{1}{2}(1 + 8) = \frac{9}{2} = 4.5$$

$$r_3(4.5) = \delta(4.5)$$

Como

$a_3 = -\frac{2}{9} < 0$ sustituimos 1 por 4.5 y el índice 5 por 3

Asi:

$r_3(4.5) = \delta(4.5)$ y $r_2(8) = \delta(8)$

Entonces

$$a_3 = -\frac{2}{9} \leq 0 \qquad a_2 = \frac{1}{5} \geq 0$$

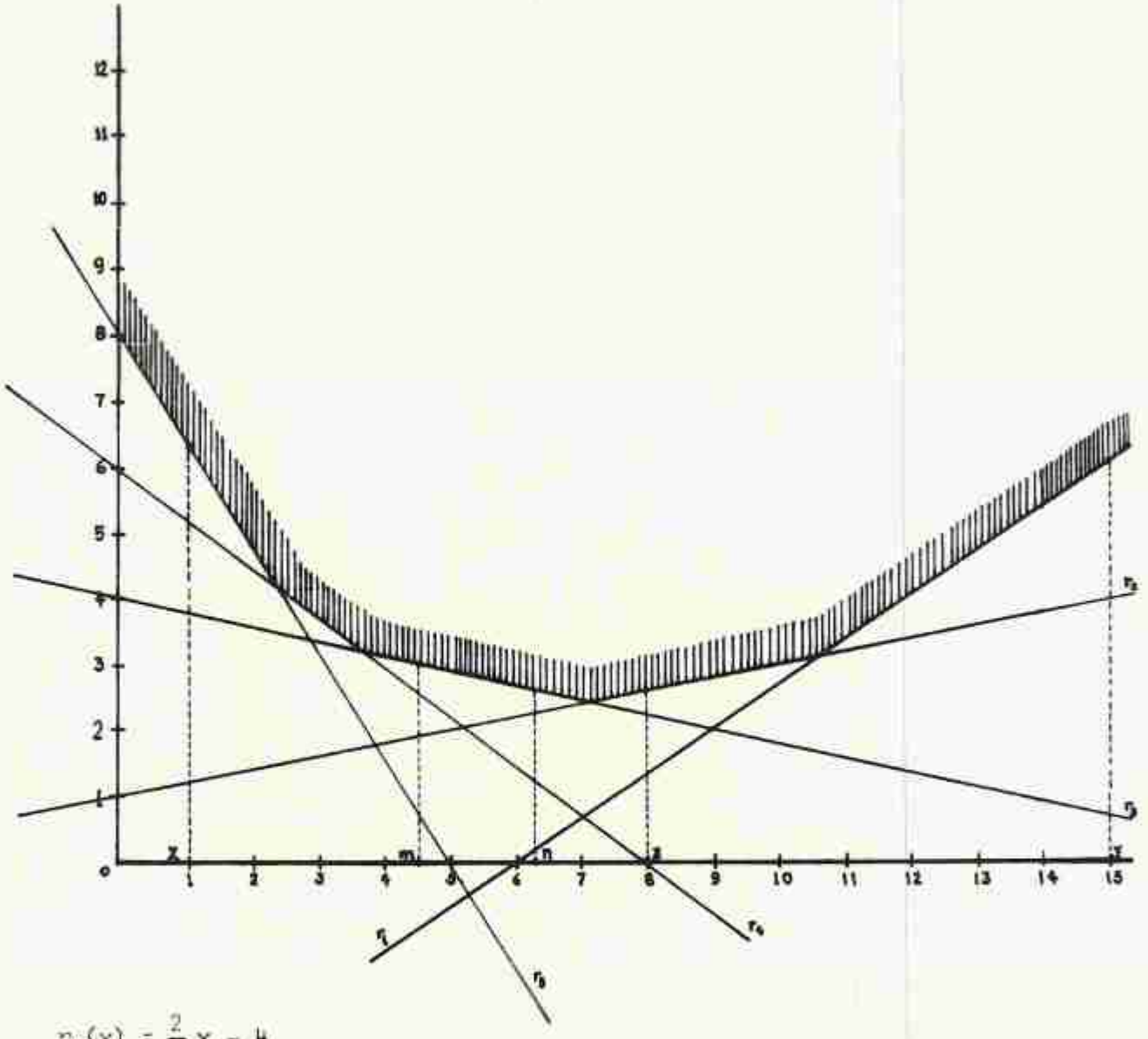
Como $a_3 \neq 0$ entonces 4.5 no es solución

$a_2 \neq 0$ entonces 8 no es solución

Pongamos

$$n = \frac{1}{2}(4.5 + 8) = \frac{12.5}{2} = 6.25$$

Ahora tenemos los puntos 6.25 y 8. Como ya no podemos sustituir valores ni índices, se deduce que la solución 1 está entre 6.25 y 8, y entre los residuales r_2 y r_3.



$r_1(x) = \frac{2}{3}x - 4$

$r2(x) = \frac{1}{5}x + 1$

$r3(x) = -\frac{2}{9}x + 4$

$r4(x) = -\frac{3}{13}x + 6$

$r_5(x) = -\frac{8}{5}x + 8$

PROBLEMAS

1. Localizar los puntos mínimos para las siguientes funciones:
(Cada una ilustra un fenómeno diferente).

a) $\max \{2 - 2x, 1 - \frac{1}{2} x\}$

b) $\max \{|2 - 2x|, |1 - \frac{1}{2} x|\}$

c) $\max \{|2 - 2x|, |1 - \frac{1}{2} x|, 1\}$

d) $\max \{|2 - 2x|, |1 - \frac{1}{2} x|, 1, x\}$

3. CARACTERIZACION DE LA SOLUCION

Consideremos nuevamente un sistema de ecuaciones lineales

$$r_i(x) = \sum_{j=1}^{n} A_j^i x_j - b_i = 0 \quad (i = 1,\ldots,m)$$

y dos funciones que dependen de él:

$$\delta(x) = \max_i \; r_i(x)$$

$$\Delta(x) = \max_i \; |r_i(x)|$$

Nuestra meta es desarrollar métodos para localizar puntos mínimos de las funciones δ y Δ. Primero es necesario descubrir qué propiedades distinguen a las soluciones de todos los otros puntos x. Denotemos la primera fila de la matriz (A_j^i) por A^i. Entonces

$$r_i(x) = \langle A^i, x \rangle - b_i$$

Teorema de Caracterización. (A)

Un punto z es un punto mínimo de la función δ si y sólo si el origen está situado en la cápsula convexa del conjunto

$$\{A^i / r_i(z) = \delta(z)\}$$

PRUEBA:

"$\Rightarrow$" Supongamos que z no es punto mínimo de δ. Entonces para algún vector h, $\delta(z - h) < \delta(z)$. Suponga que

$$M = \{i / r_i(z) = \delta(z)\}$$

Entonces para $i \in M$ tenemos:

$$r_i(z - h) \leq \delta(z - h) < \delta(z) = r_i(z) \qquad y$$

$$\langle A^i, z - h \rangle - b_i < \langle A^i, z \rangle - b_i$$

Así

$$\langle A^i, h \rangle > 0 \qquad (i \in M).$$

Por el teorema de las desigualdades lineales (T.3 del Apéndice) el origen no está situado en la cápsula convexa del conjunto

$$\{A^i / i \in M\}$$

"$\Leftarrow$" Suponga que cero no está situado en la cápsula convexa de

$$\{A^i / i \in M\} .$$

Por el citado teorema existe un h tal que $\langle A^i, h \rangle > 0$ para $i \in M$.

Por consiguiente el número

$$a = \min_{i \in M} \langle A^i, h \rangle$$

es positivo. Los residuales $r_i(z)$ para $i \in M$, decrecerán en la dirección $-h$, porque para $\lambda > 0$,

$$r_i(z - \lambda h) = r_i(z) - \lambda \langle A^i, h \rangle \leq \delta(z) - \lambda\alpha$$

Los residuales $r_i(z)$, para $i \notin M$, son menores que $\delta(z)$ y por la continuidad permanecen asi en un vecindario de z. De esta forma hay puntos cerca de z produciendo valores inferiores de δ. Los detalles de este argumento pueden ser ordenados como sigue: Supongamos que

$$\beta = \max_{i \in M} r_i(z) \qquad y \qquad \gamma = \min_{1 \leq i \leq m} \langle A^i, h \rangle.$$

Entonces para $i \notin M$

$$r_i(z - \lambda h) = r_i(z) - \lambda \langle A^i, h \rangle \leq \beta - \lambda\gamma$$

Podemos hacer todos los residuales menores que

$$C = \frac{1}{2} \left[\beta + \delta(z) \right],$$

digamos, seleccionando $\lambda > 0$ en tal forma que

$$\beta - \lambda\gamma < C \qquad y \qquad \delta(z) - \lambda\alpha < C$$

Para la función Δ se tiene un teorema similar: Nosotros lo damos sin prueba. Es conveniente usar la función Sgn, la cual es definida por

$$\operatorname{Sgn} x = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Teorema de Caracterización. (B)

Dado un punto $z \in \mathbb{R}_n$, supongamos

$$\sigma_i = \text{Sgn}\ r_i(z) \quad y \quad M = \{i/|r_i(z)| = \Delta(z)\}.$$

El punto z minimiza a Δ si y sólo si el origen de $\mathbb{R}_n$ está situado en la cápsula convexa del conjunto

$$\{\sigma_i A^i / i \in M\}$$

Llegamos ahora a la formulación n-dimensional de un caso que ya ha sido percibido para $n = 1$. Si volvemos al bosquejo de $\Delta(C)$, observaremos que en el punto mínimo, un cierto par de funciones residuales - determina el gráfico de Δ. El resultado preciso para un espacio n-dimensional es como sigue:

TEOREMA:

Si z es un punto mínimo de la función

$$\delta(x) = \max_{1 \le i \le m} r_i(x),$$

entonces z es un punto mínimo de

$$\max_{i \in J} r_i(x)$$

donde J es un cierto subconjunto de $\{1, \ldots, m\}$ abarcando al máximo $n + 1$ índices.

PRUEBA:

Mediante el teorema de caracterización (A), sabemos que el origen - de $\mathbb{R}_n$ está situado en la cápsula convexa del conjunto $\{A^i / i \in M\}$, donde

$$M = \{i/r_i(z) = \delta(z)\}.$$

Si M contiene $n + 1$ ó menos elementos, hagamos $J = M$. De otra forma, por el teorema de Caratheódory (T.4 del Apéndice), seleccionamos un subconjunto J de M teniendo al máximo $n + 1$ elementos, tal que

$$0 \in \mathcal{H}\{A^i / i \in J\} .$$

Por el teorema de Caracterización, z es también un punto mínimo de-

$$\max_{i \in J} r_i(x).$$

Para la función Δ, un resultado similar es válido y toma la forma - siguiente:

TEOREMA:

Toda solución minimax del sistema

$$\sum_{j=1}^{n} A_j^i x_j = b_i \quad (i = 1, \ldots, m > n)$$

es una solución minimax de un subsistema apropiado comprendiendo $n + 1$ ecuaciones.

Una vez que tal "subsistema apropiado" es conocido, es relativamente fácil obtener su solución minimax. En varios métodos prácticos, para computar las soluciones minimax de sistemas de ecuaciones el consumo - principal del esfuerzo está en localizar este subsistema apropiado. Pospondremos la discusión de este problema para ver primero como obtener - la solución minimax de $n + 1$ ecuaciones con n incognitas.

PROBLEMAS

1.- Determinar si el punto z = (1,1) es un punto mínimo de la función

$$\delta(x) = \max \{(x_1 + 2x2 - 4), (-x_1 + 2x2 - 3),$$

$$(-x_1 - x2 + 1), (x_1 - x2 - 1)\} .$$

Si no lo es, determinar un h tal que

$$\delta(z - h) < \delta$$

2.- Determinar si el punto z = (2,1) es un punto mínimo de la función

$$\Delta(x) = \max \{|3x_1 + x2 - 4|, |6x_1 - x2 + 5|,$$

$$|x_1 + x_2 + 2|, |-x_1 + 2x2 - 5|\}$$

Si no lo es, determinar un h tal que

$$\Delta(z - h) < \Delta(z).$$

Entonces haga el mismo problema para el punto z = (-1, 3).

4. EL CASO ESPECIAL, m = n+1

En esta sección consideramos el problema de computar la solución (o - soluciones) minimax de un sistema de n+1 ecuaciones con n incognitas:

$$r_i(x) = \sum_{j=1}^{n} A_j^i x_j - b_i = \langle A^i, x \rangle - b_i = 0 (i=1,\ldots,n+1)$$

Probablemente la forma más satisfactoria de resolver tal sistema es - el método de La Vallée Poussin. Suponga que mediante algunos medios somos- capaces de descubrir un punto x, signos $\sigma_i = \pm 1$, y un número ε tal que-

(1) $$r_i(x) = \sigma_i \varepsilon \qquad (i = 1,\ldots, n+1)$$

(2) $$0 \in \mathcal{H} \{\sigma_1 A^1, \ldots, \sigma_{n+1} A^{n+1}\}$$

Entonces afirmamos que x es una solución minimax del sistema. En efecto, por (1) tenemos $|r_i(x)| = |\varepsilon|$.

Entonces mediante el teorema de Caracterización (B) y por la propiedad (2) anterior, x es una solución. Por lo tanto nos pondremos a garantizar las condiciones (1) y (2) anteriores.

Primero encontraremos una solución no trivial de las ecuaciones lineales

$$\sum_{i=1}^{n+1} \lambda_i A^i = 0.$$

Esto es posible porque los n+1 vectores $A^1, \ldots, A^{n+1}$, siendo elementos de un espacio n-dimensional, son necesariamente linealmente dependientes. Si definimos $\sigma_i = 1$ cuando $\lambda_i \geq 0$ y $\sigma_i = -1$ cuando $\lambda_i < 0$, entonces la condición (2) ya está asegurada porque

$$0 = \Sigma(\sigma_i\lambda_i)(\sigma_i A^i).$$

Para completar la discusión asumimos que la matriz es de rango n. Asi algún conjunto de n de sus filas es linealmente independiente, y numerando de nuevo las filas si es necesario, podemos tomar este conjunto para ser $\{A^1,\ldots,A^n\}$. Ahora si la condición (1) debe ser encontrada entonces

$$\langle A^i, x\rangle - b_i = \varepsilon\,\sigma_i$$

Multiplicando esta ecuación por λ_i y sumando para $i = 1,\ldots,n$ - produce

$$\langle \Sigma\,\lambda_i A^i, x\rangle - \Sigma\,\lambda_i b_i = \varepsilon\,\Sigma\,\sigma_i\lambda_i.$$

En vista de lo que ya conocemos, esto se reduce a

$$-\Sigma\,\lambda_i b_i = \varepsilon\,\Sigma|\lambda_i|$$

y esta ecuación puede ser tomada como la definición de ε ya que

$$\Sigma\,|\lambda_i| > 0$$

Queda ser demostrada que con esta definición de ε la ecuación (1) es realmente consistente. Si nosotros dejamos fuera la ecuación correspondiente a $i = n + 1$ el sistema resultante puede ser resuelto para una x única- debido a nuestra presunción de que $\{A^1,\ldots,A^n\}$ es linealmente independiente. En esta forma tenemos $r_i(x) = \sigma_i\,\varepsilon$ para $i = 1,\ldots,n$. Pero - nosotros ya hemos visto que

$$\sum_{i=1}^{n+1} \lambda_i r_i(x) = \varepsilon \sum_{i=1}^{n+1} \sigma_i\lambda_i.$$

Por lo tanto

$$\lambda_{n+1}\, r_{n+1}(x) = \varepsilon\,\sigma_{n+1}\,\lambda_{n+1}$$

Si $\lambda_{n+1} = 0$, entonces, la ecuación

$$\Sigma\lambda_i A^i = 0$$

representaría una dependencia lineal entre $A_i, \ldots, A_n$, lo cual es imposible.

Consecuentemente

$$\lambda_{n+1} \neq 0 \quad y \quad r_{n+1}(x) = \varepsilon\sigma_{n+1}$$

Esencialmente el mismo método puede ser descrito con el uso de determinantes. Suponga, para tomar un ejemplo concreto que estamos confrontando con el sistema

$$x_1 - x_2 = 7$$
$$2x1 + 3x2 = 5$$
$$3x_1 + x_2 = -1$$

Las condiciones (1) en este caso se leeran

$$x_1 - x_2 - \sigma_1\varepsilon = 7$$
$$2x1 + 3x2 - \sigma_2\varepsilon = 5$$
$$3x_1 + x_2 - \sigma_3\varepsilon = -1$$

Por la regla de Cramer,

$$\varepsilon = \begin{vmatrix} 1 & -1 & 7 \\ 2 & 3 & 5 \\ 3 & 1 & -1 \end{vmatrix} \div \begin{vmatrix} 1 & -1 & -\sigma_1 \\ 2 & 3 & -\sigma_2 \\ 3 & 1 & -\sigma_3 \end{vmatrix} = \frac{-74}{7\sigma1 + 4\sigma2 - 5\sigma3}$$

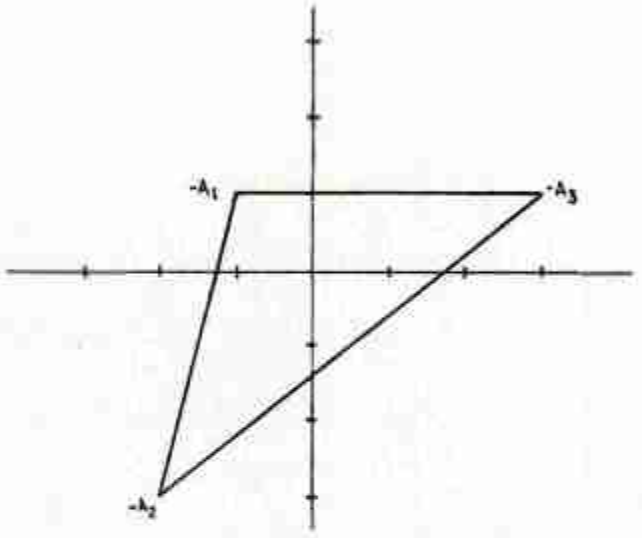
Para que ε tenga valor positivo mínimo, debemos tomar

$$\sigma_1 = -1, \quad \sigma_2 = -1 \quad y \quad \sigma_3 = 1,$$

con lo cual se obtiene $\varepsilon = \frac{7}{8}$. Conociendo $\sigma_1, \sigma_2, \sigma_3$ y ε, fácilmente calculamos

$$x_1 = \frac{3}{2} \quad y \quad x_2 = -\frac{7}{8}.$$

Observe que 0 está situado en la cápsula convexa de los vectores $\sigma_i A^i$:



Los dos métodos que acabamos de describir para obtener los signos σ_i no son necesariamente los mismos desde el punto de vista computacional pero teoricamente son los mismos. En efecto, en el segundo método los σ_i fueron seleccionados como los signos de los números -7, -4, 5, y éstos últimos pueden servir como los números λ_i del primer método.

Así, si los λ_i son tomados igual a sus propios cofactores en el determinante

$$\begin{vmatrix} 1 & -1 & \lambda_1 \\ 2 & 3 & \lambda_2 \\ 3 & 1 & \lambda_3 \end{vmatrix}$$

entonces $\lambda 1\begin{pmatrix}1\\-1\end{pmatrix} + \lambda 2\begin{pmatrix}2\\3\end{pmatrix} + \lambda 3\begin{pmatrix}3\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$

ya que si los elementos de una columna de un determinante son multiplicados por los cofactores de una columna diferente y sumados, el resultado es cero.

Un hecho notable que emerge de la discusión precedente es que para un (n+1)xn sistema de ecuaciones con rango n, una vez los signos σ_i son conocidos, la solución minimax puede ser obtenida resolviendo un conjunto consistente de n + 1 ecuaciones lineales con n + 1 incognitas. En problemas de este tipo que surgen de la aproximación de funciones continuas en un intervalo, los signos pueden ser usualmente determinados fácilmente sin resolver ninguna ecuación lineal. En el caso general que estamos discutiendo aquí, sin embargo, los signos necesitan una computación separada de aquel para las x_i y ε.

Es interesante que los signos de los residuales son los mismos para las soluciones de cuadrados mínimos y la solución minimax cuando m=n+1. (Nosotros continuamos asumiendo que nuestra matriz es de rango n). Puesto que esto es verdadero para una clase amplia de normas, además de la norma de cuadrados mínimos, nosotros nos quedaremos brevemente en esta aserción. La discusión necesita de la noción de un hiperplano. Un hiperplano en un espacio vectorial E es un conjunto de puntos de la forma

$$H = \{x \in E / f(x) = c\}$$

donde c es una constante y f es una función lineal no cero de valores reales. (En el caso de un espacio de Banach requerimos también que f sea continua). Asi en el espacio R_{n+1} un hiperplano consiste de todas

las (n + 1)-uplas u para las cuales $\langle u,f \rangle = c$, siendo f una (n+1)-upla fija y c una constante. Ahora cuando resolvemos aproximadamente un sistema de ecuaciones

$$\sum_{j=1}^{n} A_j^i x_j = b_i \quad (i=1,\ldots, n+1)$$

estamos intentando hacer el vector

$$r = \Sigma\, x_j\, A_j - b$$

tan pequeño como sea posible en la norma. Los puntos de la formar, estan situados en un hiperplano, un hecho que estableceemos formalmente como sigue.

LEMA

Si el conjunto de vectores $\{A_1,\ldots, A_n\}$ es independiente en R_{n+1} y si b e s un elemento fijo de R_{n+1}, entonces el conjunto de puntos

$$\{ \sum_{j=1}^{n} x_j A_j - b / x_j \text{ real}\}$$

es un hiperplano.

PRUEBA:

Por el teorema Gram-Schmidt (T.5 del Apéndice) podemos seleccionar un vector no cero u ortogonal para $A_1,\ldots, A_n$. Ponga $c = -\langle u,b \rangle$

Si

$$r = \Sigma\, x_j\, A_j - b$$

entonces claramente $\langle u,r \rangle = c$. Por otra parte, suponga que

$$\langle u, r \rangle = c.$$

Puesto que $\{A1, \ldots, A_n, u\}$ es una base para $\mathbb{R}_{n+1}$, una ecuación

$$r + b = \Sigma x_j A_j + x_0 u$$

es posible al tomar el producto interno de ambos lados de esta ecuación con u produce

$$0 = x_0 \langle u, u \rangle, \quad \text{de donde} \quad x_0 = 0$$

Otra prueba de este teorema puede ser basado en la siguiente idea:

Por el teorema de existencia de las mejores aproximaciones (T.1 - Apéndice) existe allí un vector y para el cual la norma euclidiana

$$||r(y)||$$

es un mínimo. Podemos mostrar que $r(y)$ es ortogonal a cada vector A_j. En realidad, de la definición de

$$y, \quad ||r(y) - \lambda A_j||^2 \geq ||r(y)||^2$$

de donde

$$-2\lambda \langle r(y), A_j \rangle + \lambda^2 ||A_j||^2 \geq 0$$

Si esto es verdadero para todos los λ reales entonces

$$\langle r(y), A_j \rangle = 0.$$

Así para toda x,

$$\langle r(y), r(x) \rangle = -\langle r(y), b \rangle$$

Si $r(y) \neq 0$, entonces esto muestra que todos los vectores $r(x)$ están situados en un cierto hiperplano, probando la mitad del teorema. - Puesto que la ecuación anterior es verdadera cuando $x = y$ encontramos que

$$-\langle r(y), b \rangle = ||r(y)||_2$$

En el siguiente teorema necesitamos el concepto de una norma monóto na en $\mathbb{R}_n$. Tal norma tiene la propiedad que

$$||x|| \leq ||y||$$

siempre que los vectores

$$x = [x_1, \ldots, x_n] \quad y \quad y = [y_1, \ldots, y_n]$$

estan relacionados por las desigualdades

$$|x_i| \leq |y_i| \quad \text{para} \quad i = 1, \ldots, n.$$

Todas las normas

$$||x||_p = \sqrt[p]{\Sigma\ |x_i|^p} \qquad (1 \leq p \leq \infty)$$

tiene esta propiedad. Pero la norma en $\mathbb{R}_2$

$$||x|| = \max \{2|x_1 + x_2|, |x_1 - x_2|\}$$

es un ejemplo simple de la norma que no es monótona.

TEOREMA:

Sea H un hiperplano en $\mathbb{R}_n$. Los puntos de H que minimizan dos normas monótonas diferentes tiene componentes que concuerdan en signos (o pueden ser seleccionados en el caso de la no unicidad).

PRUEBA:

Si $0 \in H$, el teorema es trivial, y por consiguiente asumimos lo contrario.

Sea $H = \{x \in \mathbb{R}_n / \langle x, u \rangle = c\}$

Cambiando u a - u si es necesario, podemos asumir que $c > 0$. Sea x un punto de H que minimiza una norma monótona $||x||$. Poniendo

$$x_i' = |x_i| \operatorname{Sgn} u_i$$

obtenemos un punto x' cuyas componentes concuerdan en signo con aquellos - de u. Así

$$\langle x', u \rangle \geq \langle x, u \rangle = c > 0,$$

y el número

$$\theta = c / \langle x', u \rangle$$

está situado en el intervalo $(0,1]$. Por la monotonicidad de la norma,

$$||\theta x'|| \leq ||x'|| \leq ||x||$$

Puesto que

$\langle \theta x', u \rangle = c$, $\theta x'$ está situada en H y minimiza la - norma.

Del teorema precedente vemos que los signos σ_i necesitados para resolver el problema minimax para n+1 ecuaciones con n incognitas pueden - ser obtenidos resolviendo el sistema en el sentido de los cuadrados mínimos y usando para σ_i el signo del primer residual de los cuadrados mínimos. (Esta afirmación permanece verdadera para otras normas monótonas - además de la Euclidiana, pero es de menos significancia práctica). Llevando esta idea un paso más lejos, encontramos que el número ε en el sistema (1) puede también ser determinado. Nosotros resumimos nuestros descubrimientos como sigue:

TEOREMA

Sea y la solución de cuadrados mínimos de un sistema de n+1 ecuaciones lineales con n incognitas, $r_i(x) = 0$.

Asuma que el sistema es de rango n. Entonces la solución minimax es la solución exacta del sistema

$$r_i(x) = \sigma_i \varepsilon, \text{ donde } \sigma_i = \operatorname{Sgn} r_i(y) \quad y \quad \varepsilon = \Sigma r_i^2 (y) / \Sigma |r_i(y)|$$

PRUEBA:

Por el lema anterior los puntos

$$r(x) = \Sigma\, x_j A_j - b$$

completa un hiperplano H en E_{n+1}. Realmente las observaciones que siguieron al lema mostraron que

$$H = \{z / \langle z, r(y)\rangle = \langle r(y), r(y)\rangle\}$$

Si definimos un punto z con componentes $z_i = \sigma_i \cdot \varepsilon$, entonces $z \in H$ porque

$$\langle z, r(y)\rangle = \Sigma\, z_i r_i(y) = \varepsilon\, \Sigma\, \sigma_i r_i(y) = \varepsilon\, \Sigma\, |r_i(y)| = \Sigma\, r_i^2(y)$$

Por otra parte, ningún punto de H está más cerca de cero que z en la norma Tchebycheff. Ya que si

$||u||_T < ||z||_T$, entonces $\langle u, r(y)\rangle = \Sigma\, u_i r_i(y) \leq \max |u_i| \Sigma\, |r_i(y)|$

$$< \max |z_i|\ \Sigma\, |r_i(y)|$$

$$= \varepsilon\, \Sigma\, |r_i(y)|$$

$$= \Sigma\, r_i^2(y),$$

de tal modo que u no está situada en H.

Como una ilustración de este teorema consideramos el sistema

$$x_1 - x_2 = 7$$

$$2x_1 + 3x_2 = 5$$

$$3x_1 + x_2 = -1$$

el cual fue resuelto anteriormente, usando el método de La Vallée Poussin. Para obtener la solución de cuadrados mínimos podemos minimizar la función

$$\phi(x_1, x_2) = (x_1 - x_2 - 7)^2 + (2x_1 + 3x_2 - 5)^2 + (3x_1 + x_2 + 1)^2$$

Haciendo las derivadas parciales de ϕ igual a cero, tenemos

$$14x_1 + 8x2 = 14$$

$$8x1 + 11x2 = 7$$

de donde $x_1 = \frac{49}{45}$ y $x_2 = -\frac{7}{45}$. El vector residual correspondiente a esta solución de cuadrados mínimos llega a ser

$$(-\frac{259}{45}, -\frac{148}{45}, \frac{185}{45})$$

y el número ε llega a ser $\frac{37}{8}$. La solución minimax de nuestro sistema es por consiguiente la solución exacta del sistema.

$$x_1 - x_2 - 7 = -\frac{37}{8}$$

$$2x_1 + 3x_2 - 5 = -\frac{37}{8}$$

$$3x_1 + x_2 + 1 = \frac{37}{8}$$

Por supuesto, el vector residual minimax será

$$(-\frac{37}{8}, -\frac{37}{8}, \frac{37}{8})$$

La solución minimax misma es

$$(\frac{3}{2}, -\frac{7}{8})$$

PROBLEMA

1. Un punto Tchebycheff en el hiperplano $(x/\langle u,x\rangle = c)$ es cualquier punto para el cual $||x||_T$ es un mínimo.

   Muestre que el punto cuyas componentes son $x_i = \varepsilon \operatorname{Sgn} u_i$ con

   $\varepsilon = \frac{c}{\sum |u_i|}$ es tal punto.

5. ALGORITMO DE POLYA

En esta sección consideramos el primero de varios métodos para resolver el sistema inconsistente general de ecuaciones

$$\sum_{j=1}^{n} A_j{}^i x_j = b_i \quad (i = 1,\ldots,m)$$

en la minimax o sentido Tchebycheff. Se señaló en la sec . 3 que para este propósito podemos descartar todos excepto n + 1 de las ecuaciones dadas, las n + 1 que son retenidas generalmente no son conocidas al principio. Ahora el algoritmo a ser descrito descansa sobre una idea de Pólya y puede ser usado para determinar este conjunto crucial de n + 1 ecuaciones.

Recordemos la familia de normas que fueron definidas por la ecuación

$$||v||_p = \left(\sum_{i=1}^{m} |v_i|^p\right)^{1/p} \quad (p \geq 1)$$

para cualquier vector $v = (v_1, \ldots, v_m)$ de E_m . Para un vector fijo v, los números $||v||_p$ convergen a

$$||v||_T = \max |v_i| \quad \text{cuando } p \to \infty$$

Realmente la convergencia es monotonamente hacia abajo. Denotemos ahora por $v^{(p)}$ el punto de la forma $\Sigma x_j A_j - b$ para el cual $||v||_p$ es un mínimo y similarmente $v^{(T)}$. Mediante esta definición y la monotonicidad justamente mencionada tenemos

$$(1) \qquad ||v^{(T)}||_T \leq ||v^{(p)}||_T \leq ||v^{(p)}||_p \leq ||v^{(T)}||_p .$$

Dejando $p \to \infty$, tenemos

$$||v^{(T)}||_p + ||v^{(T)}||_T$$

y consecuentemente

$$||v^{(p)}||_T \to ||v^{(T)}||_T .$$

Así para valores grandes de p, $v^{(p)}$ es un buen sustituto para $v^{(T)}$. Observe que no estamos diciendo que $v^{(p)} \to v^{(T)}$. Sin alguna aclaración posterior, esta afirmación sería sin significado de cualquier modo, porque $v^{(T)}$ no necesita ser única. Fácilmente podemos ver que si $v^{(T)}$ es único, entonces $v^{(p)} \to v^{(T)}$. En efecto, los puntos $v^{(p)}$ son acotados ya que, por ejemplo,

$$||v^{(p)}||_T \leq ||v^{(T)}||_1 .$$

Asi de la familia $\{v^{(p)}/p \geq 1\}$ podemos extraer una sucesión convergente de puntos $v^{(p_1)}, v^{(p_2)}, \ldots$ $(p_k \to \infty)$. Ahora, para el punto limite v de tal sucesión tenemos, usando la continuidad de $||.||_T$ y la desigualdad (1) anterior,

$$||v||_T = ||\lim_k v^{(p_k)}||_T = \lim_k ||v^{(p_k)}||_T = ||v^{(T)}||_T .$$

Puesto que $v^{(T)}$ es único, $v = v^{(T)}$, ya que esto es verdadero para cualquier sucesión convergente de $\{v^{(p)}\}$, tenemos $\lim v^{(p)} = v^{(T)}$. Es verdadero, pero no nos detenemos a probarlo, que $v^{(p)}$ converge a uno de los puntos $v^{(T)}$, aun si el último no es único.

Un algoritmo que es sugerido por estas observaciones, consiste simplemente en calcular para valores sucesivamente mayores de p, los puntos $v^{(p)}$, y extrapolar numéricamente para el vector limite. En la práctica podemos hacerlo mucho mejor usando los puntos $v^{(p)}$ solamente para-

determinar alguna información cualitativa acerca de $v^{(T)}$, y entonces - resolviendo precisamente para $v^{(T)}$. En realidad, para p grande, podemos esperar (debido a los resultados de la sec . 3) que $v^{(p)}$ excibirá n+1 componentes aproximadamente iguales y de magnitudes máximas. Esto nos dice cuales n+1 ecuaciones vamos a retener del conjunto original de m y también cuales deberían ser los signos σ_i de los residuales. Nosotros entonces estaríamos en una posición para usar los métodos de la - sección precedente. El método falla, si por accidente hay más de n+1 - residuales máximos iguales en magnitud en la solución Tchebycheff.

A través de la discusión nos hemos referido a los vectores residuales $v \in R_m$ más bien que a los vectores coeficiente $x \in R_n$. Si la matriz de los coeficientes A_j^i es de rango total (rango n), entonces x - es únicamente determinada mediante v por vía de las ecuaciones

$$v_i = \sum_{j=1}^{n} A_j^i x_j - b_i$$

Para el trabajo numérico es conveniente minimizar la función

$$\phi(x) = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_j^i x_j - b_i \right|^{2p}$$

ya que los mínimos de $||v||_{2p}$ y $||v||_{2p}^{2p}$ ocurren en el mismo lugar. - El exponente se usa para facilitar la diferenciación. El mínimo puede - ser buscado por el método del descenso más inclinado, o resolviendo las ecuaciones algebráicas

$$\frac{\partial \phi}{\partial x} = 0 \quad \text{por el método de Newton.}$$

Como una ilustración damos un sistema de 6 ecuaciones con dos in-

cognitas y mostramos las soluciones aproximadas para

$$p = 2,\ 4,\ 6,\ 40,\ 100,\ 400$$

y la solución Tchebycheff, $p = \infty$

El ejemplo ha sido construído para indicar que el conjunto crítico de n+1 filas no puede ser disponible para los valores muy bajísimos - de p.

$$x_1 + x_2 = 3$$
$$x_1 - x_2 = 1$$
$$x_1 + 2x_2 = 7$$
$$2x1 + 4x_2 = 11.1$$
$$2x1 + x_2 = 6.9$$
$$3x_1 + x_2 = 7.2$$

	p = 2	p = 4	p = 6	p = 40	p = 100	p = 100	p = ∞
x_1	2.0741	2.0466	2.0390	2.0141	2.0055	2.0014	2.0000
x_2	1.8078	1.9207	1.9498	1.9938	1.9977	1.9994	2.0000
r_1	0.88196	0.96730	0.98887	1.0078	1.0032	1.0008	1.0000
r_2	-0.73373	-0.87411	-0.91082	-0.97955	-0.99214	-0.99809	-1.0000
r_3	-1.3102	-1.1120	-1.0613	-0.99855	-0.99911	-0.99977	-1.0000
r_4	0.27961	0.67601	0.77742	0.90290	0.90178	0.90045	0.90000
r_5	-0.94392	-0.88610	-0.87211	-0.87810	-0.89125	-0.89786	-0.90000
r_6	0.83020	0.86049	0.86692	0.83602	0.81428	0.80349	0.80000
ϕ	2.1659	1.4452	1.2610	1.0257	1.0102	1.0025	1.0000

PROBLEMAS

1. Teorema de Jensen. Para $v = (v1, \ldots, v_m)$ fijo, la función

$$||v||_p = \sqrt[p]{\sum_{i=1}^{m} |v_i|^p}$$

es una función decreciente de p.

Sugerencia:

Si $||v|| = 1$, entonces $|v_i| \leq 1$ y consecuentemente

$$|v_i|^q \leq |v_i|^p \quad \text{cuando } q \geq p.$$

Use la homogeneidad de la norma para remover la restricción

$$||v||_p = 1.$$

2. Para v fijo, $||v|| \to ||v||_T = \max |v_i|$ cuando $p \to \infty$

Sugerencia:

Si $||v||_T = 1$ entonces $1 \leq \sqrt[p]{\Sigma |v_i|^p} \leq \sqrt[p]{n} \to 1$

Use el problema 1 y la homogeneidad de la norma.

6. EL ALGORITMO DE ASCENSO

En esta sección consideramos el método general de ascenso, el cual fue ilustrado con n=1 en la sección 2 (pág. 12). Con el propósito de hacer que los cálculos continuen sin ningún problema, vamos a suponer acerca de la matriz $(A_j^{i)}$ que sus filas satisfacen un requerimiento algo fuerte de no degeneración llamada la condición de Haar: Se dice que un conjunto de vectores en un espacio n dimensional satisface la condición de Haar, si todo conjunto de n de ellos es linealmente independiente. Expresado de otra forma, cada selección de n vectores de tal conjunto es una base para un espacio n-dimensional.

Teorema de Cambio:

Sea $\{A^0, \ldots, A^{n+1}\}$ un conjunto de vectores en un espacio n-dimensional satisfaciendo la condición de Haar.

Si 0 está situado en la cápsula convexa de $\{A^0, \ldots, A^n\}$, entonces existe un índice $j \leq n$ tal que esta condición permanece verdadera cuando A_j es reemplazada por A^{n+1}.

PRUEBA:

Por hipotesis existen constantes $\theta_i \geq 0$ tal que

$$0 = \sum_{i=0}^{n} \theta_i A^i \quad y \quad \sum_{i=0}^{n} \theta_i = 1$$

La condición de Haar seria violada si cualquier θ_i fuera cero, por consiguiente $\theta_i > 0$. Nosotros podemos, por lo tanto, resolver para cualquier A_j, obteniendo

$$A^j = \sum_{\substack{i=0 \\ i \neq j}}^{n} - \frac{\theta i}{\theta_j} A^i$$

Puesto que $(A^0, \ldots, A^n)$ genera el espacio n-dimensional, es posible escribir

$$A^{n+1} = \sum_{i=0}^{n} \lambda_i A^i \quad \text{para } \lambda_i \text{ apropiado.}$$

Por consiguiente

$$0 = A^{n+1} - \sum_{i=0}^{n} \lambda_i A^i$$

$$= A^{n+1} - \lambda_j A^j - \sum_{\substack{i=0 \\ i\neq j}}^{n} \lambda_1 A^i$$

$$= A^{n+1} - \lambda_j \sum_{i=0}^{n} - \frac{\theta_i}{\theta_j} A^i - \sum_{\substack{i=0 \\ i\neq j}}^{n} \lambda_i A^i$$

$$= A^{n+1} + \sum_{\substack{i=0 \\ i\neq j}}^{n} \left(\frac{\lambda_i \theta_i}{\theta_j} - \lambda_i\right) A^i$$

Ahora si j es seleccionada de tal forma que

$$\frac{\lambda_j \theta_i}{\theta_j} - \lambda_i \geq 0,$$

entonces nuestra ecuación final anterior expresará al cero como una - combinación lineal no negativa de $A^0, \ldots, A^{n+1}$, donde A^j no aparece, lo cual sería suficiente para probar que cero está en la cápsula convexa de estos puntos. Nuestro requerimiento de j es que

$\frac{\lambda_j}{\theta_j} \geq \frac{\lambda_i}{\theta_i}$, en otras palabras nosotros debemos seleccionar j de tal forma que

$\frac{\lambda_j}{\theta_j}$ sea la más grande de estas proporciones. (Este índice j es único, porque si hubieran dos proporciones mayo-

res $\frac{\lambda_i}{\theta_j}$, entonces uno de los coeficientes en la ecuación anterior desaparecería, contradiciendo la condición de Haar).

Ahora procedemos a una descripción del algoritmo. Buscamos un punto donde la función

$$\Delta(x) = \max_{1\leq i\leq m} |r_i(x)| = \max_{1\leq i\leq m} |\langle A^i, x\rangle - b_i|$$

obtiene su valor mínimo. Se asume que la condición de Haar se satisface mediante el conjunto de vectores $\{A^1, \ldots, A^m\}$. La idea básica de este algoritmo es calcular las soluciones minimax de una colución de subsistemas, cada uno abarcando n+1 ecuaciones. Por el teorema en pág. 23, la solución de uno de estos subsistemas es el punto buscado. Por otro lado, puede haber no obstante un número finito de estos subsistemas y esta simple observación será la base para una prueba de que el algoritmo es efectivo.

En cada ciclo computado, tendremos un conjunto de n+1 índices

$$J = \{i_0, \ldots, i_n\}$$

y un vector de signos $\sigma = \{\sigma_0, \ldots, \sigma_n\}$ tal que

(1) $$0 \in \mathcal{CC}\{\sigma_0 A^{io}, \ldots, \sigma_n A^{in}\}$$

Entonces resolvemos el siguiente sistema de n+1 ecuaciones lineales para determinar un vector $y = (y_1, \ldots, y_n)$ y un número e:

(2) $$\sigma_j r_{ij}(y) = e \quad (j = 0, \ldots, n)$$

Para asegurar que $e > 0$, podemos cambiar los signos de todos los σ_j sin perder la propiedad (1). Como ya hemos visto en la pág. 23, las

condiciones (1) y (2) implican que y es una solución minimax del sistema.

$$\langle A^{ij}, y \rangle = b_{ij} \quad (j = 0, \ldots, n)$$

Si $e = \Delta(y)$, entonces por el teorema de caracterización (B), y es una solución minimax del sistema original de m ecuaciones. En el caso contrario, existe ahí por lo menos un índice α (no en J) tal que $|r_\alpha(y)| > e$. Ordinariamente tomaríamos α de modo que $|r_\alpha(y)| = \Delta(y)$, pero esto no es necesario. Supongamos ahora que $\mu = \text{Sgn } r_\alpha(y)$. Usando el teorema de cambio, sustituimos uno de los vectores

$$\sigma_o A^{io}, \ldots, \sigma_n A^{in}$$

por μA^α en tal forma que el origen permanezca en la cápsula convexa del conjunto, [propiedad (1)]. La situación presentada ahora es la misma que aquella en el principio y procedemos como antes.

La necesidad de calcular y Δ e nos obliga a hacer alguna presunción acerca de los datos dados, una presunción conveniente es que la matriz

$$\begin{bmatrix} \sigma_o & A_1^{io} \ldots A_n^{io} \\ \ldots\ldots\ldots\ldots\ldots \\ \sigma_n & A_1^{in} \ldots A_n^{in} \end{bmatrix}$$

sea no singular. Encontraremos casos, más tarde, cuando esta condición pueda ser verificada por consideraciones apriori.

Las computaciones del algoritmo terminan solamente cuando $e = \Delta(y)$, y esta ecuación significa que y es una solución. Ya hemos observado que solamente un número finito de conjuntos J existen. Todo lo que falta por

ser demostrado entonces, es que las computaciones no "forman ciclos", es decir, que no regresan infinitamente muchas veces al mismo subconjunto J Esta prueba será realizada mostrando que el número e es una función de J el cual aumenta estrictamente de paso a paso. Para este fin supongamos por simplicidad de notación que en un cierto paso J es $\{1,\ldots, n+1\}$ y que α es $n+2$. Supongamos, ademas, que en el próximo paso, J llega a ser $J' = \{2,\ldots, n+2\}$.

Sea y' Λ e' los valores de y Λ e correspondientes al nuevo conjunto J'. Por la selección de α,

$$|r_2(y)| < |r_{n+2}(y)| ,$$

mientras que

$$|r_2(y')| = |r_{n+2}(y')| .$$

Por consiguiente $y - y' \neq 0$. Puesto que

$$\langle \sigma_i A^i, y - y' \rangle = \sigma_1 r_i(y) - \sigma_1 r_1(y')$$

$$= e - e' \quad \text{para} \quad i = 2,\ldots, n+1,$$

las condiciones de Haar implican que $e - e' \neq 0$. Ahora

$$\langle \sigma_{n+2} A^{n+2}, y - y' \rangle = \sigma_{n+2} r_{n+2}(y) - \sigma_{n+2} r_{n+2}(y') > e - e' .$$

Si $e - e' > 0$, entonces $\langle \sigma_i A^i, y - y' \rangle > 0$ para $i = 2,\ldots,n+2$ contradiciendo el hecho que

$$0 \in \mathcal{K}\{\sigma_i A^i / 2 \leq i \leq n + 2\}$$

[Recuerde el teorema de las desigualdades lineales T.3 de Apéndice].

Por lo tanto podemos concluir que $e - e' < 0$, como debió ser probado.

Un punto pequeño puede necesitar clasificación: ¿Cómo son determina

dos el conjunto inicial J y el vector σ ? J puede ser tomado arbitrariamente, y entonces podemos encontrar una solución no trivial a la ecuación

$$\sum_{j=0}^{n} a_j A^{ij} = 0$$

poniendo entonces

$$\sigma_j = \operatorname{Sgn} a_j .$$

Un arreglo conveniente de las computaciones es como sigue. Observe primero que las ecuaciones (2), es decir,

$$e = \sigma_j r_{ij}(y) = \sigma_j [\langle A^{ij}, y \rangle - b_{ij}],$$

pueden ser escritas como

$$-\sigma_j e + \langle A^{ij}, y \rangle = b_{ij}$$

y por eso en notación matricial como

$$\begin{bmatrix} \sigma_o A_1^{io} & \cdots & A_n^{io} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \sigma_n A_1^{in} & \cdots & A_n^{in} \end{bmatrix} \begin{bmatrix} -e \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} b_{io} \\ \cdot \\ \cdot \\ \cdot \\ b_{in} \end{bmatrix}$$

Si asumimos que la matriz A_J de la izquierda tiene un inverso $C = (C_j^i)$, entonces podemos escribir

$$\begin{bmatrix} -e \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} C_o^o & \cdots & C_n^o \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ C_o^n & & C_n^n \end{bmatrix} \begin{bmatrix} b_{io} \\ \cdot \\ \cdot \\ \cdot \\ b_{in} \end{bmatrix}$$

Puesto que C es el inverso de A_y, tenemos

$$\begin{bmatrix} C_o^o \ldots C_n^o \\ \ldots\ldots\ldots \\ C_o^n \ldots C_n^n \end{bmatrix} \begin{bmatrix} \sigma_o A_1^{io} \ldots A_n^{io} \\ \ldots\ldots\ldots\ldots\ldots \\ \sigma_n A_1^{in} \ldots A_n^{in} \end{bmatrix} = \begin{bmatrix} 1 \ldots 0 \\ \ldots\ldots\ldots \\ 0 \ldots 1 \end{bmatrix}$$

De esto es evidente que

$$\sum_{j=0}^{n} \sigma_j C_j^o = 1 \qquad y \qquad \sum_{j=0}^{n} C_j^o A^{ij} = 0.$$

Así los números $\sigma_j C_j^o$ son los coeficientes necesitados para expresar el hecho de que cero está situado en la cápsula convexa de los puntos $\sigma_j A^{ij}$. Estos coeficientes entran en los cálculos relacionados al teorema de Cambio. De la prueba de ese teorema vemos que debemos expresar μA^{α} como una combinación lineal de

$$\sigma_o A^{io}, \ldots \sigma_n A^{in} .$$

Si ponemos

$$A^{\alpha} = \sum_{j=0}^{n} \lambda_j A^{ij} ,$$

entonces los coeficientes λ_j pueden ser obtenidos resolviendo la matriz

$$(\lambda_o, \ldots, \lambda_n) \begin{bmatrix} \sigma_o A_1^{io} \ldots A_n^{io} \\ \ldots\ldots\ldots\ldots\ldots \\ \sigma_n A_1^{in} \ldots A_n^{in} \end{bmatrix} = (\mu, A_1^{\alpha}, \ldots, A_n^{\alpha})$$

de la cual la solución es

$$(\lambda_o, \ldots, \lambda_n) = (\mu, A_1^{\alpha}, \ldots, A_n^{\alpha}) \begin{bmatrix} C_o^o \ldots C_n^o \\ \ldots\ldots\ldots \\ C_o^n \ldots C_n^n \end{bmatrix}$$

Puesto que $A^a = \sum_{j=0}^{n} (\mu\sigma_j\lambda_j)(\sigma_j A^{ij})$, las proporciones a ser computadas en el teorema de cambio son

$$\mu\sigma_j\lambda_j/\sigma_j C_j^o = \mu\lambda_j/C_j^o .$$

El número β es seleccionado como el índice del mayor de estas proporciones.

Antes de resumir el algoritmo en un diagrama de flujo, deberíamos observar que avanzando de un ciclo del cálculo al siguiente, solamente una fila de la matriz A_j cambia.

El efecto en C puede ser predecido por el uso del siguiente teorema.

TEOREMA.

Sea A una matriz no singular y $C_1, \ldots, C_n$ las columnas de su inverso. Sea $\overline{A}$ la matriz obtenida mediante el reemplazo de la fila β-esima de A por un vector v. Si

$$\lambda = \langle v, C_\beta \rangle \neq 0,$$

entonces $\overline{A}$ es no singular y las columnas de su inverso son dadas por las reglas

$$\overline{C}_\beta = \lambda^{-1} C_\beta \quad y \quad \overline{C}_j = C_j - \langle v, C_j \rangle \overline{C}_\beta \quad (j \neq \beta)$$

PRUEBA

Para verificar que $\overline{AC} = I$, computamos el producto interno de $\overline{A}^i$ (la i-esima fila de $\overline{A}$) con $\overline{C}_j$. Hay cuatro casos. En el caso 1, $i = \beta$ y $j = \beta$. Entonces

$$\langle \overline{A}_\beta, \overline{C}_\beta \rangle = \langle v, \lambda^{-1} C_\beta \rangle = \lambda^{-1} \langle v, C_\beta \rangle = 1.$$

En el caso 2, $i \neq \beta$ y $j = \beta$. Entonces

$$\langle \bar{A}^i, \bar{C}_\beta \rangle = \langle A^i, \lambda^{-1} c_\beta \rangle = \lambda^{-1} \langle A^i, c_\beta \rangle = 0.$$

En el caso 3, $i = \beta$ y $j \neq \beta$. Entonces

$$\langle \bar{A}^i, \bar{C}_j \rangle = \langle v, C_j \rangle - \langle v, C_j \rangle \bar{C}_\beta = \langle v, c_j \rangle - \langle v, c_j \rangle \lambda^{-1} \langle v, C_\beta \rangle = 0.$$

En el Caso 4, $i \neq \beta$ y $j \neq \beta$. Entonces

$$\langle \bar{A}^i, \bar{C}_j \rangle = \langle A^i, C_j - \langle v, C_j \rangle \bar{C}_\beta \rangle$$

$$= \langle A^i, C_j \rangle - \langle v, C_j \rangle \lambda^{-1} \langle A^i, C_\beta \rangle = \langle A^i, C_j \rangle = \delta_{ij}$$

En el presente algoritmo reemplazamos la fila

$$(\sigma_\beta, A1^{i\beta}, \ldots, A_n^{i\beta}) \text{ por una nueva fila}$$

$$(\mu, A_1^\alpha, \ldots, A_n^\alpha)$$

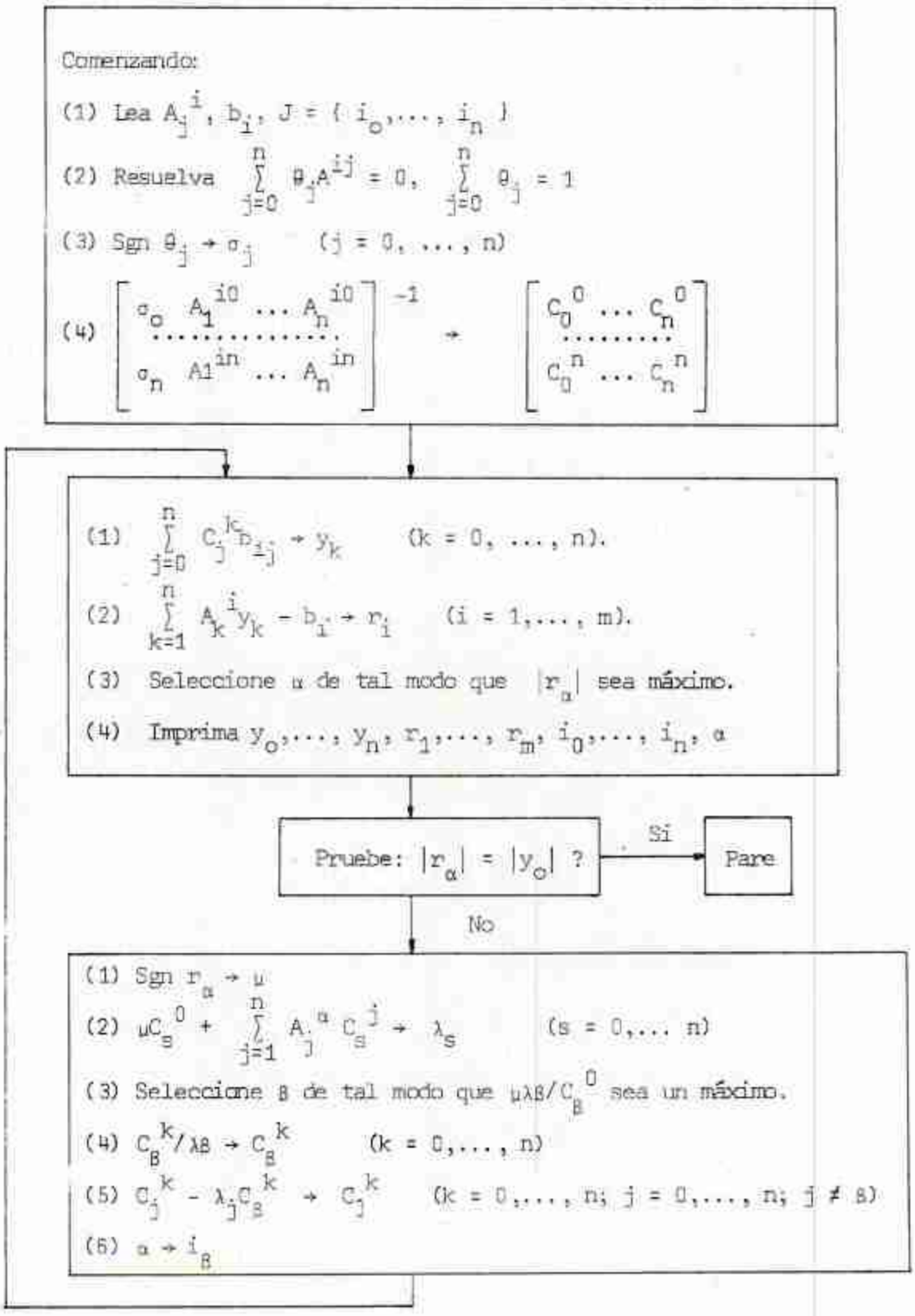
Necesitaremos el número λ el cual es el producto interno de la nueva fila con β-ésima columna de C, es decir,

$$\mu C^o + \sum_{j=0}^{n} A_j^\alpha C_\beta^j .$$

Pero este es el número λ_β previamente computado. En el diagrama de - flujo es Cuestión de simplicidad denotar con y_o el número -e. También, no hemos dado los detalles para los pasos (2) y (4) en la primera caja, siendo estos problemas standar en el algebra lineal. Finalmente, la escritura corta $x \to w$ denota que la cantidad x sustituye la cantidad w o que la cantidad de x está almacenada en la memoria en la celda marcada w.

El diagrama de flujo para minimizar la función

$\Delta(x) = \max_{1 \leq i \leq m} | \langle A^i, x \rangle - b_i |$ mediante el método del ascenso es mostrado así:


Comenzando:
(1) Lea A_j^i, b_i, $J = \{ i_0, \ldots, i_n \}$
(2) Resuelva $\sum_{j=0}^{n} \theta_j A^{ij} = 0$, $\sum_{j=0}^{n} \theta_j = 1$
(3) Sgn $\theta_j \to \sigma_j$ $(j = 0, \ldots, n)$
(4) $\begin{bmatrix} \sigma_0 & A_1^{i0} & \ldots & A_n^{i0} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_n & A1^{in} & \ldots & A_n^{in} \end{bmatrix}^{-1} \to \begin{bmatrix} C_0^0 & \ldots & C_n^0 \\ \ldots & \ldots & \ldots \\ C_0^n & \ldots & C_n^n \end{bmatrix}$
(1) $\sum_{j=0}^{n} C_j^k b_{i_j} \to y_k$ $(k = 0, \ldots, n)$.
(2) $\sum_{k=1}^{n} A_k^i y_k - b_i \to r_i$ $(i = 1, \ldots, m)$.
(3) Seleccione α de tal modo que $|r_\alpha|$ sea máximo.
(4) Imprima $y_0, \ldots, y_n, r_1, \ldots, r_m, i_0, \ldots, i_n, \alpha$
Pruebe: $|r_\alpha| = |y_0|$?
Si
Pare
No
(1) Sgn $r_\alpha \to \mu$
(2) $\mu C_s^0 + \sum_{j=1}^{n} A_j^\alpha C_s^j \to \lambda_s$ $(s = 0, \ldots n)$
(3) Seleccione β de tal modo que $\mu\lambda\beta / C_\beta^0$ sea un máximo.
(4) $C_\beta^k / \lambda\beta \to C_\beta^k$ $(k = 0, \ldots, n)$
(5) $C_j^k - \lambda_j C_\beta^k \to C_j^k$ $(k = 0, \ldots, n;\ j = 0, \ldots, n;\ j \neq \beta)$
(6) $\alpha \to i_\beta$

7. EL ALGORITMO DESCENDENTE.

En esta sección consideramos el método descendente general, el - cual fue ilustrado con n=1 en la pág. 8 . Por variedad, la discusión - es dirigida hacia la minimización de la función

$$\delta(x) = \max_{1 \leq i \leq m} \{\langle A^i, x \rangle - b_i\}$$

La manera en la cual esto incluye el problema Tchebycheff se discutió en la pág. 4. La idea general del método es avanzar hacia abajo de vértice a vértice en la hiper-superficie del espacio (n+1) dimensional cuya ecuación es

$$z = \delta(x)$$

Una técnica especial es usada para llegar a un primer vértice.

Sea x^o cualquier vector inicial. Numerando de nuevo los residuales, por conveniencia en la descripción podemos asumir que

$$\delta(x^o) = r_1(x^o) = r_2(x^o) = \ldots = r_k(x^o) > r_{k+i}(x^o) \quad (i \geq 1)$$

Buscamos una dirección a mover desde x^o en la cual los residuales $r1, \ldots, r_k$ disminuyan con una velocidad igual. Avanzamos en esta dirección hasta que una función residual (k+1)-ésimo surja para encontrar el primer k.

Aplicando de nuevo esta técnica en a lo sumo n pasos llegamos a un vértice, donde ocurren n+1 residuales iguales al máximo. Ahora la velocidad de cambio en r_i a medida que nos movemos de x^o en la dirección y se - ve fácilmente que es el número $\langle A^i, y \rangle$. En efecto

$$\frac{d}{d\lambda} r_i (x^o + \lambda y) = \frac{d}{d\lambda} \left[\langle A^i, x^o + \lambda y \rangle - b_i \right] = \langle A^i, y \rangle$$

Por consiguiente una dirección y en la cual los residuales $r_1, \ldots, r_k$ permanecen iguales disminuyendo a una velocidad común puede ser determinada resolviendo las ecuaciones

(1) $$\langle A^i, y \rangle = -1 \quad (i = 1, \ldots, k)$$

Si asumimos las condiciones de Haar (es decir, cada conjunto de n vectores A^i es independiente), entonces esta condición sobre y es *fácilmente encontrada siempre que* $k \leq n$.

En realidad, podemos tomar y de la forma

$$y = \sum_{j=1}^{k} C_j A^j$$

siendo entonces las ecuaciones

$$\sum_{j=1}^{k} C_j \langle A^i, A^j \rangle = -1 \quad (i = 1, \ldots, k)$$

No nos detendremos a probar la no singularidad de la matriz de Gram, $\langle A^i, A^j \rangle$, pero la prueba se da en el apéndice T.6

Habiendo determinado un vector y con las propiedades (1), tomamos nuestro siguiente punto que es de la forma

$$x^1 = x^o + \lambda y$$

donde λ es el menor coeficiente positivo para el cual ocurren k+1 residuales iguales al máximo.

Asumamos ahora que un punto x^o es conocido donde los residuales iguales al máximo ocurren, digamos,

$$\delta(x^\circ) = r_1(x^\circ) = \ldots = r_{n+1}(x^\circ) > r_{n+i}(x^\circ), \qquad (i > 1)$$

Es concebible que x° es una solución. Este será el caso si y sólo - si el sistema de desigualdades lineales

$$\langle A^i, y \rangle < 0 \qquad (i = 1, \ldots, n+1)$$

es inconsistente. Si este sistema es consistente entonces mediante el - teorema de las desigualdades lineales, (T.3 de Apéndice) cero no está en la cápsula convexa de

$$\{A^1, \ldots, A^{n+1}\}$$

Consecuentementesi resolvemos las siguientes ecuaciones para

$\theta_1, \ldots, \theta_{n+1}$,

$$\sum_{i=1}^{n+1} \theta_i A^i = 0 \qquad \sum_{i=1}^{n+1} \theta_i = 1$$

entonces por lo menos un coeficiente θ_i será negativo. Ahora saliendo - desde el vértice asociado con el punto x° hay un número de "ángulos" los cuales son variedades lineales de una dimensión a lo largo de los cuales los n residuales permanecen iguales entre si. Sin embargo no todos estos ángulos están situados realmente en la superficie

$$z = \delta(x)$$

Para cada conjunto de n índices seleccionados de $\{1, \ldots, n+1\}$ hay- un ángulo a lo largo del cual aquellos n residuales son iguales. El resi- dual que queda, digamos r_j, generalmente cambiará en una velocidad diferente.

La dirección de éste ángulo será por lo tanto obtenido resolviendo- un sistema tal como la siguiente para el vector y^j.

$$\langle A^i, y^j \rangle = -1 \quad (i = 1, \ldots, n+1,\ i \neq j)$$

En la dirección y^j, el residual r_j cambiará en una proporción $\langle A^j, y^j \rangle$, y este puede ser mayor o menor que -1. Si $\langle A^j, y^j \rangle < -1$, entonces el ángulo correspondiente está situado en la superficie. Para computar $\langle A^j, y^j \rangle$ escribimos

$$0 = \langle 0, y^j \rangle = \sum_{i=1}^{n+1} \theta_i \langle A^i, y^j \rangle = \sum_{\substack{i=1 \\ i \neq j}}^{n+1} -\theta_i + \theta_j \langle A^j, y^j \rangle$$

Asi

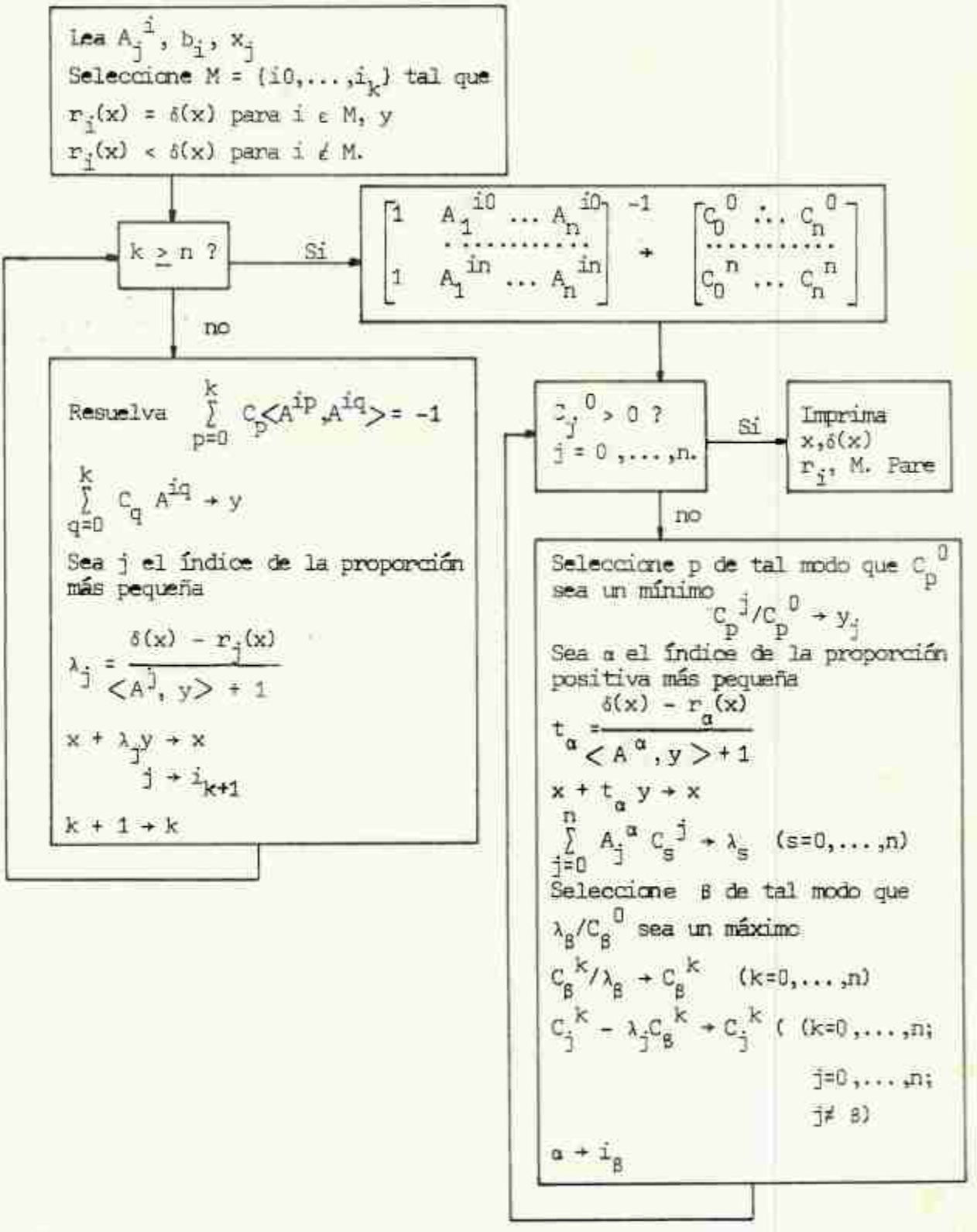
$$\langle A^j, y^j \rangle = \frac{1}{\theta_j} \sum_{i=1}^{n+1} \theta_i = \frac{1 - \theta_j}{\theta_j} = \frac{1}{\theta_j} - 1$$

Ciertamente una de estas proporciones será menor que -1 porque al menos uno de los θ_j es negativo. Si j es tal índice el siguiente puntoes de la forma $x^o + \lambda y^j$, donde tomamos λ como el menor coeficiente positivo para el cual ocurren n+1 residuales iguales al máximo. Desde esta etapa, la computación puede ser simplificada usando el teorema de la pág. 49.

El diagrama de flujo para minimizar la función

$$\delta(x) = \max_{1 \leq i \leq m} \{ \langle A^i, x \rangle - b_i \}$$

por el método descendente se muestra aquí:


Lea A_j^i, b_i, x_j
Seleccione $M = \{i0,\ldots,i_k\}$ tal que
$r_i(x) = \delta(x)$ para $i \in M$, y
$r_i(x) < \delta(x)$ para $i \notin M$.
$k \geq n$?
Si
no
$\begin{bmatrix} 1 & A_1^{i0} \ldots A_n^{i0} \\ \cdots & \cdots \\ 1 & A_1^{in} \ldots A_n^{in} \end{bmatrix}^{-1} \rightarrow \begin{bmatrix} C_0^0 \ldots C_n^0 \\ \cdots \\ C_0^n \ldots C_n^n \end{bmatrix}$
Resuelva $\sum_{p=0}^{k} C_p \langle A^{ip}, A^{iq} \rangle = -1$
$\sum_{q=0}^{k} C_q A^{iq} \rightarrow y$
Sea j el índice de la proporción más pequeña
$\lambda_j = \frac{\delta(x) - r_j(x)}{\langle A^j, y \rangle + 1}$
$x + \lambda_j y \rightarrow x$
$j \rightarrow i_{k+1}$
$k + 1 \rightarrow k$
$C_j^0 > 0$?
$j = 0,\ldots,n.$
Si
Imprima
$x, \delta(x)$
r_i, M. Pare
no
Seleccione p de tal modo que C_p^0 sea un mínimo
$C_p^j / C_p^0 \rightarrow y_j$
Sea α el índice de la proporción positiva más pequeña
$t_\alpha = \frac{\delta(x) - r_\alpha(x)}{\langle A^\alpha, y \rangle + 1}$
$x + t_\alpha y \rightarrow x$
$\sum_{j=0}^{n} A_j^\alpha C_s^j \rightarrow \lambda_s \quad (s=0,\ldots,n)$
Seleccione β de tal modo que
$\lambda_\beta / C_\beta^0$ sea un máximo
$C_\beta^k / \lambda_\beta \rightarrow C_\beta^k \quad (k=0,\ldots,n)$
$C_j^k - \lambda_j C_\beta^k \rightarrow C_j^k$ ($(k=0,\ldots,n;$
$j=0,\ldots,n;$
$j \neq \beta)$
$\alpha \rightarrow i_\beta$

CAPITULO II

APROXIMACIÓN TCHEBYCHEFF POR POLINOMIOS

1. INTRODUCCION

En este capítulo consideramos el problema de aproximar una función contínua f definida en un intervalo $[a,b]$ por un polinomio

$$P(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_o .$$

Nuestro interés se centra en las aproximaciones que minimizan las expresiones de la forma

$$(1) \qquad \max_{a \le x \le b} | f(x) - P(x)|$$

ó

$$(2) \qquad \max_{1 \le i \le m} | f(x_i) - P(x_i)|$$

Algo de la discusión puede ser dado, sin embargo, para el problema de aproximación más general en el cual los monomios $1, x, x^2, \dots, x^n$ son reemplazados por otras funciones fijas $g_o, g_1, \dots, g_n$. En conexión con ésta podemos hablar de polinomios generalizados, entendiendo como tales, funciones de la forma

$$\sum_{i=0}^{n} c_i g_i$$

Asi trataremos de abarcar en nuestra teoría un problema de aproximación de la forma extravagante

$$f(x) \approx c_1 \log x + c_2 \cos x + c_3 e^x + c_4 (x - 2)^{-1}$$

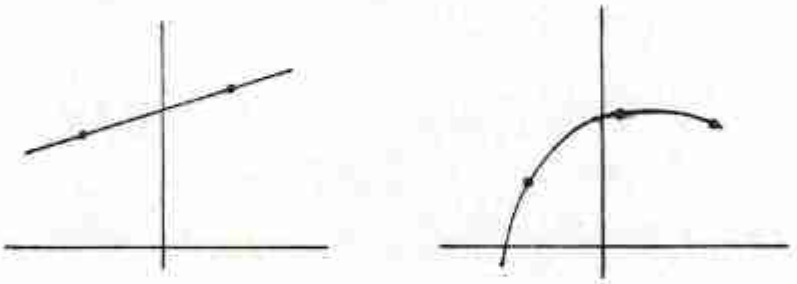
Un caso muy especial de la teoría ocurre cuando en la expresión (2)

anterior el número de puntos se toma igual a n+1. En éste caso el problema de aproximación tiene una solución explicita inmediata. Discutimos este tópico, la interpolación, primero; las técnicas desarrolladas aquí serán útiles en otros tipos de aproximación.

2. INTERPOLACION.

Sabemos que una línea recta que tiene la ecuación $y = ax + b$ puede trazarse a través de cualquier par de puntos con abscisas distintas. Similarmente una parábola

$y = ax^2 + bx + c$ puede trazarse a través de cualquier tripleta de puntos con abscisas distintas.



El resultado general a lo largo de éstas líneas fácilmente supuesto, toma la siguiente forma.

TEOREMA DE INTERPOLACION

Existe un polinomio único de grado menor o igual que n el cual toma valores dados en n+1 puntos distintos.

PRUEBA 1.

Sean $x_o,\ldots, x_n$ los puntos y $y_o,\ldots, y_n$ los valores prescritos. Buscamos un polinomio P tal que $P(x_i) = y_i$ $(i = 0,\ldots, n)$. Ya que el polinomio es de grado $\leq n$, puede ser expresado como

$$P(x) = \sum_{j=0}^{n} c_j x^j.$$ Por consiguiente nuestro requerimiento se lee ahora

$$\sum_{j=0}^{n} c_j x_i^j = y_i \quad (i = 0,\ldots,n).$$

Escrita en forma de matriz, esta se convierte en:

$$\begin{bmatrix} 1 & x_o & x_o^2 \ldots & x_o^n \\ . & . & . \ldots . & \\ . & . & . \ldots . & \\ . & . & . \ldots . & \\ 1 & x_n & x_n^2 \ldots & x_n^n \end{bmatrix} \begin{bmatrix} c_o \\ . \\ . \\ . \\ c_n \end{bmatrix} = \begin{bmatrix} y_o \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

En esta ecuación por supuesto, las c son las incognitas, mientrasque la matriz x y el lado derecho son conocidos. Esta ecuación tiene una solución única porque la matriz coeficiente es no singular. El determinante de esta matriz, conocido como determinante de Vandermonde, tiene el valor

$$D = \prod_{o \leq j < i \leq n} (x_i - x_j)$$

El lado derecho de ésta fórmula denota el producto de todos los factores $(x_i - x_j)$ para los cuales el par (i, j) satisface $0 \leq j < i \leq n$. De ésta fórmula para D, es claro que $D \neq 0$ si y sólo si los puntos x_i son distintos.

PRUEBA 2.

Nuestro polinomio podía ser escrito inmediatamente si existieran polinomios ℓ_i de grado $\leq n$ con la propiedad $\ell_i(x_j) = \delta_{ij}$. En efecto - escribiríamos

$$P(x) = \Sigma\, y_i \ell_i(x), \text{ de donde}$$

$$P(x_j) = \sum_i y_i \ell_i(x_j) = \sum_i y_i\, \delta_{ij} = y_j$$

Una reflexión de momento muestra que ℓ_i es

$$\ell_i(x) = \prod_{\substack{j=o \\ j\neq i}}^{n} \frac{x - x_j}{x_i - x_j}$$

Otra forma de definir ℓ_i es comenzar con

$$W(x) = \prod_{j=o}^{n} (x - x_j)$$

Entonces

$$\ell_i(x) = a_i \frac{W(x)}{x - x_i}$$

Ahora el requerimiento $\ell_i(x_i) = 1$ conduce, vía regla de L'Hospital a $1 = a_i W'(x_i)$

La fórmula que resulta se conoce como la fórmula de interpolación de Lagrange:

$$P(x) = \sum_{i=o}^{n} y_i \ell_i(x), \quad \ell_i(x) = \prod_{\substack{j=o \\ j\neq i}}^{n} \frac{x - x_j}{x_i - x_j} = \frac{W(x)}{(x - x_i)W'(x_i)}$$

Falta por probar la unicidad de P. Suponga entonces que P y Q son

dos polinomios de grado $\leq n$ los cuales tienen la propiedad $Q(x_i) = P(x_i) = y_i$. Entonces $P - Q$ es un polinomio de grado $\leq n$ el cual desaparece en los $n+1$ puntos distintos x_i. Por lo tanto $P - Q = 0$.

PRUEBA 3.

Intentemos determinar el polinomio deseado en la forma tal

$$P(x) = a_o + a_1(x-x_o) + a_2(x-x_o)(x-x1)+\ldots + a_n(x-x_o) \ldots (x-x_{n-1})$$

Haciendo $x = x_o$, encontramos que $P(x_o) = a_o$. Ya que $P(x_o)$ es prescrito, igual y_o, debemos tomar $a_o = y_o$. con a_o ahora conocido, procedemos a poner $x = x_1$, obteniendo $P(x1) = a_o + a_1(x_1 - x_o)$. Resolviendo para los coeficientes uno por uno de esta manera produce:

$$a_o = y_o$$

$$a_1 = \frac{y_1 - a_o}{x_1 - x_o}$$

$$a_2 = \frac{y_2 - a_o - a_1(x_2 - x_o)}{(x_2 - x_o)(x_2 - x_1)}$$

$$a_3 = \frac{y_3 - a_o - a_1(x_3-x_o) - a_2(x_3-x_o)(x_3-x_1)}{(x_3 - x_o)(x_3 - x_1)(x_3 - x_2)}$$

etc.

La existencia de P entonces, resulta del hecho que en esta forma los denominadores no desaparecen. No nos detenemos a dar una prueba de unicidad basada en la fórmula anterior.

Cada una de las tres pruebas acabadas de dar sugieren un procedimiento numérico diferente para obtener el polinomio de interpolación. Para-

ilustrar esto determinemos el cuadrático

$$P(x) = c_o + c_1 x + c_2 x^2$$

pasando a través de los 3 puntos (1,2), (2,-1), (3,1), usando cada método en turno.

En el primer método resolvemos las ecuaciones

$$c_o + c_1 + c_2 = 2$$

$$c_o + 2c_1 + 4c_2 = -1$$

$$c_o + 3c_1 + 9c_2 = 1$$

obteniendo como resultado

$$P(x) = 10 - \frac{21}{2} x + \frac{5}{2} x^2$$

Usando la fórmula de Lagrange, P tiene la forma

$$P(x) = (x - 2)(x - 3) + (x - 1)(x - 3) + \frac{1}{2} (x - 1)(x - 2).$$

Finalmente, usando el método de la 3a. prueba, obtenemos P en la forma:

$$P(x) = 2 - 3(x - 1) + \frac{5}{2} (x - 1)(x - 2)$$

Volvemos ahora a la pregunta de fijar el proceso de interpolación como un instrumento de aproximación. Al principio de este capítulo, dos medidas relacionadas de la discrepancia entre dos funciones f y P fueron introducidas, a saber,

$$\max_{a \leq x \leq b} |f(x) - P(x)| \quad \text{y} \quad \max_{1 \leq i \leq m} |f(x_i) - P(x_i)|$$

El polinomio P de grado $\leq n$ el cual interpola a f en n+1 puntos x_i, claramente resuelve el problema de minimizar la 2a. expresión cuando -

$m = n + 1$. Preguntamos, ¿será también pequeña la 1a. expresión cuando se selecciona P en esta forma? La respuesta ciertamente es "NO" si la conducta de f entre los puntos de interpolación no es de algún modo controlado. Resulta que tal control es posible para las funciones que poseen $n + 1$ derivadas contínuas.

TEOREMA.

Si f posee n derivadas continuas en $[a,b]$, si P es el polinomio de grado $< n$ el cual interpola a f en n nodos x_i en $[a,b]$, y si $W(x) = \Pi(x - x_i)$, entonces, en términos de la norma Tchebycheff,

$$||f - P|| \leq \frac{1}{n!} ||f^{(n)}|| \; ||W||$$

PRUEBA

Probaremos algo más, es decir que a cada x en $[a,b]$ corresponde un $\xi \in (a,b)$ tal que

$$f(x) - P(x) = \frac{1}{n!} f^{(n)}(\xi) \; W(x) \tag{1}$$

Esta fórmula es obvia si x es uno de los nodos. De otra forma, ponemos $\phi = f - P - \lambda W$ donde λ se selecciona para hacer $\phi(x) = 0$. Está claro que ϕ desaparece también en todos los nodos x_i. Así ϕ desaparece en al menos $n + 1$ puntos de $[a,b]$. Mediante el teorema de Rolle, ϕ' desaparece por lo menos una vez entre cualquier par de ceros de ϕ y así desaparece en por lo menos n puntos. Continuando este argumento vemos que $\phi^{(n)}$ tiene por lo menos una raíz en el intervalo, digamos en el punto ξ. Pero $\phi^{(n)} = f^{(n)} - \lambda n!$ ya que P es un polinomio de grado $< n$

y $W(z) = z^n + \ldots$

Así $f^{(n)}(\xi) = \lambda n!$ Puesto que el valor de λ es

$$[f(x) - P(x)]/W(x),$$

la prueba es completa.

Respondamos ahora a la pregunta surgida en una forma natural por el teorema anterior. ¿Cómo podemos situar los nodos para optimizar el error de acotación? Ya que los nodos entran en ésta fórmula solamente en la función W, debemos intentar minimizar la norma de W.

TEOREMA

La norma uniforme de $W(x) = \prod_{i=1}^{n} (x - x_i)$ es minimizada en $[-1,1]$ cuando $x_i = \cos[(2_i - 1)\,\pi/2n]$.

PRUEBA

Se sabe que cos n θ puede ser expresado en la forma

$$\sum_{k=0}^{n} a_k \cos^k\theta$$

con coeficientes apropiados a_k, donde el coeficiente a_n es dado por 2^{n-1}. Poniendo $T_n(x) = \sum_{k=0}^{n} a_k x^k$, tenemos $T(\cos\theta) = \cos n\theta$. Las n raíces de T_n son por lo tanto los puntos x_i dados anteriormente. El polinomio $W = 2^{1-n}T_n$ es de la forma anteriormente contemplada ya que su coeficiente a_n es la unidad. El máximo de $|W(x)|$ en $[-1,1]$ ocurre entonces en los puntos $y_i = \cos i\pi / n$ ya que $T_n(y_i) = \cos i\pi = (-1)^i$. Ahora, si es posible, sea V otro polinomio de la misma forma como W pa-

ra el cual $||V|| < ||W||$. Entonces $V(y_o) < W(y_o)$, $V(y_1) > W(y_1)$, etc.. de los cuales resulta que W - V debe desaparecer por lo menos una vez en cada intervalo (y_1, y_o), $(y_2, y_1)\cdots$ para un total de n veces. Pero esto no es posible porque ambos V y W tienen coeficiente a_n unitario, y la diferencia de ellos es por lo tanto de grado < n.

TEOREMA(Interpolación de Hermite)

Existe un polinomio único P de grado $\leq 2n-1$ tal que P y su derivada P' toma los valores prescritos en n puntos.

PRUEBA

Deje que las condiciones en P sean tales que $P(x_i) = y_i$ y $P'(x_i) = y'_i$ para i = 1,..., n. Como en el caso de interpolación ordinaria nuestro polinomio puede ser escrito en una forma explícita (fórmula de interpolación de Hermite)

$$P(x) = \sum_{i=1}^{n} \left[y_i A_i(x) + y'_i B_i(x) \right]$$ donde A_i y B_i son polinomios de grado $\leq 2n-1$ con las propiedades $A_i(x_j) = \delta_{ij}$, $B_i(x_j) = 0$, $A'_i(x_j) = 0$ y $B'_i(x_j) = \delta_{ij}$. En términos de la función $\ell_i(x)$, A_i y B_i toman la forma

$$A_i(x) = \left[1 - 2(x - x_i)\, \ell'_i(x_i) \right] \ell_i^2(x)$$

$$B_i(x) = (x - x_i)\ell^2{}_i(x).$$

Se deja al lector el verificar usando la ecuación $\ell_i(x_i) = \delta_{ij}$ que el polinomio P así definido tiene las propiedades interpolares deseadas.

A fin de probar la unicidad de P, nosotros suponemos lo contrario, que existe otro polinomio Q, de grado $\leq 2n-1$ teniendo las propiedades

$$Q(x_i) = y_i \quad y \quad Q'(x_i) = y'_i$$

Luego P - Q es un polinomio que tiene raices de al menos, multiplicidad 2 en cada punto x_i, puesto que $(P - Q)'(x_i) = 0$. Debido a P - Q es de grado $\leq 2n - 1$ deberá ser cero.

PROBLEMAS

Polinomios de Tchebycheff.

1. a) Pruebe que $\text{Cos}(n+1)\,\theta = 2\,\text{Cos}\,\theta\ \text{Cos}\,n\,\theta - \text{Cos}\,(n-1)\theta$. La fórmula $\text{Cos}\,(A \pm B) = \text{Cos}\,A\ \text{Cos}\,B \mp \text{sen}\ A \text{ sen}\ B$ será de gran utilidad.

   b) Probar por inducción que el $\text{Cos}\,n\,\theta$ se puede expresar en la forma $T_n(\text{Cos}\,\theta) = \sum_{k=0}^{n} a_{nk}\,\text{Cos}^k\,\theta$, y que el $\text{Cos}^n\theta$ es expresable en la forma $\sum_{k=0}^{n} b_{nk}\,\text{Cos}\,k\,\theta$.

2. Probar la siguiente propiedad de T_n

$$(1 - x^2)\,T''_n(x) - x\,T'_n(x) + n^2\,T_n(x) = 0$$

3. EL TEOREMA DE WEIERSTRASS.

En la sección precedente vimos como construir un polinomio P_n de grado $\leq n$ el cual concordó con una función dada f en ciertos n+1 puntos. Queda una cuestión de especulación: si $P_n(x) \to f(x)$ cuando $n \to \infty$ para todos los puntos x. Parecería razonable, en vista del teorema de pág. 63, que para f, función uniforme (digamos una que posee derivadas de todos los órdenes), $||P_n - f|| \to 0$. Sin embargo esto no debe ser esperado en general. Específicamente si tomamos $f(x) = (x^2 + 1)^{-1}$ y computamos los polinomios de interpolación Lagrangianos P_n para nodos igualmente espaciados en $[-5, 5]$, encontramos que $||P_n - f||$ llega a ser arbitrariamente grande. Esta situación es algo sorprendente, ya que la función f es regular en todos los puntos además de $\pm i$. Este ejemplo es debido a Runge.

Otro ejemplo es debido a Bernstein: Los polinomios P_n de grado n que interpolan la función $f(x) = |x|$ en n+1 puntos igualmente espaciados en $[-1, 1]$ convergen a $f(x)$ solamente en +1, 0 y -1.

Podría sospecharse que el espaciamiento igual de los nodos fue en cierto modo culpable de este estado desafortunado de cosas. En efecto veremos después que un agrupamiento de los nodos cerca de las extremidades del intervalo es usualmente aconsejable. Sin embargo, Faber mostró que no importa como son prescritos los nodos para la interpolación, habrá alguna función contínua cuyos polinomios de interpolación fallen para la convergencia uniforme.

Contra el fundamento de estos resultados "negativos", el teorema-

de Weierstrass parece el más notable. De acuerdo a dicho teorema, existe, alguna sucesión de polinomios que convergen uniformemente a una función contínua prescrita, en un intervalo acotado cerrado. Los ejemplos anteriores indican que tales sucesiones de polinomios no pueden ser obtenidos por la interpolación en un conjunto fijo de nodos. Por otra parte, necesita ser apenas señalado que las series de Taylor no están disponibles tampoco para este propósito, excepto para una clase muy pequeña de funciones contínuas. Esta clase no incluye aun la función elemental $|x|$. Para apreciar las dificultades involucradas, consideremos esta función en el intervalo $[-1,1]$. Comenzamos con las series de Taylor para $\sqrt{1-z}$, la cual converge uniformemente para $0 \leq z \leq 1$.

$$\sqrt{1-z} = 1 - \frac{1}{2}z - \frac{1}{2\cdot 4}z^2 - \frac{1\cdot 3}{2\cdot 4\cdot 6}z^3 - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 8}z^4 - \ldots$$

Entonces sustituimos z por $1 - x^2$ para obtener

$$|x| = \sqrt{x^2} = \sqrt{1-(1-x^{2)}}$$

$$= 1 - \frac{1}{2}(1-x^{2)} - \frac{1}{2\cdot 4}(1-x^2)^2 - \frac{1\cdot 3}{2\cdot 4\cdot 6}(1-x^2)^3 - \ldots$$

Debe ser especialmente notado que esta no es una serie de Taylor en x. La prueba de Lebesgue del teorema Weierstrass usa esta serie en una manera esencial.

TEOREMA DE APROXIMACION DE WEIERSTRASS.

Sea f una función contínua definida en $[a,b]$. Para cada $\varepsilon > 0$ corresponde un polinomio P tal que $||f - P|| < \varepsilon$

Asi

$$|f(x) - P(x)| < \varepsilon \quad \text{para todo } x \in [a,b]$$

Vamos a derivar este teorema como una consecuencia de otro teorema más poderoso. Por la forma de introducir este teorema consideremos en el bosquejo la prueba dada mediante Bernstein del Teorema de Weierstrass Bernstein construyó, para una función dada $f \in C[0,1]$, una sucesión de polinomios (llamados ahora polinomios de Bernstein) $B_n f$ por medio de la fórmula

$$(1) \qquad (B_n f)(x) = \sum_{k=0}^{n} f(\tfrac{k}{n}) \binom{n}{k} x^k (1 - x)^{n-k}$$

Aquí $\binom{n}{k}$ es el coeficiente binomial $\frac{n!}{(n-k)!\ k!}$. La fórmula (1) también define para cada n un operador lineal B_n. Queremos decir mediante esto que a cada elemento f en $C[0,1]$ corresponde otro elemento $B_n f$ de $C[0,1]$ en tal forma que la condición de linealidad es encontrada.

$$(2) \qquad B_n(\alpha f + bg) = a\, B_n f + b\, B_n g$$

Se ve facilmente que los operadores B_n tienen otra propiedad expresada por la implicación

$$(3) \qquad f \geq g \Rightarrow B_n f \geq B_n g$$

Nosotros pretendemos por una desigualdad $f \geq g$ que $f(x) \geq g(x)$ para todo x (en el dominio de f). Un operador para el cual (3) es verdadero, se dice que es un operador monótono. Un examen de la prueba de Bernstein revela que el enigma de la cuestión es la verificación de que estos operadores tienen las propiedades de convergencia

(4) $B_n f \to f$ para $f(x) = 1, x, x^2$.

La conclusión de la prueba es, por supuesto, que $B_n f \to f$ para toda $f \in C[0,1]$, la convergencia aquí está en el sentido de la norma - uniforme.

Una correcta generalización de este teorema de Bernstein ha sido - dada recientemente por Bohman y por Korovkin.

Este resultado es que las propiedades (2), (3) y (4) son suficientes para que cualquier sucesión de operadores L_n tenga la propiedad $L_n f \to f$ para toda $f \in C[0,1]$.

TEOREMA DE OPERADORES MONOTONOS

Para una sucesión de operadores lineales monótonos L_n en $C[a,b]$ las condiciones siguientes son equivalentes:

i) $L_n f \to f$ (uniformemente) para toda $f \in C[a,b]$

ii) $L_n f \to f$ para las tres funciones $f(x) = 1, x, x^2$.

iii) $L_n 1 \to 1$ y $(L_n \phi_t)(t) \to 0$ uniformemente en t, donde

$\phi_t(x) \equiv (t - x)2$

PRUEBA

La implicación (i) $\Rightarrow$ (ii) es trivial

Para la prueba que (ii) $\Rightarrow$ (iii)

defina $f_i(x) = x^i$. Ya que $\phi_t(x) = t2 - 2tx + x^2$, tenemos

$\phi_t = t^2 f_o - 2t f_1 + f_2$ y $L_n \phi_t = t^2 L_n f_o - 2t L_n f_1 + L_n f_2$. Así

$$(L_n\phi_t)(t) = t2\left[(L_n f_o)(t) - 1 \right] - 2t\left[(L_n f1)(t) - t \right] + \left[(L_n f2)(t) - t2 \right]$$

$$\leq t^2 ||L_n f_o - f_o|| + |2t| \; ||L_n f_1 - f_1|| + ||L_n f_2 - f_2||$$

Puesto que t^2 y $|2t|$ son acotados en $[a,b]$, vemos que $(L_n\phi_t)(t)$ converge uniformemente a cero, probando asi (iii).

Para la prueba que (iii) $\Rightarrow$ (i)

Sea f un elemento arbitrario de $C[a,b]$. Dado $\varepsilon > 0$, obtendremos la desigualdad $||L_n f - f|| < 3\varepsilon$ para todo n suficientemente grande. Comience seleccionando $\delta > 0$ tal que

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

Ahora ponga

$\alpha = 2||f|| \; \delta^{-2}$, y sea t arbitrario pero punto fijo de $[a,b]$. Si $|t - x| < \delta$, entonces $|f(t) - f(x)| < \varepsilon$, mientras que si $|t - x| \geq \delta$, entonces

$$|f(t) - f(x)| \leq 2||f|| \leq 2||f||(t-x)^2/\delta^2 = \alpha\phi_t(x).$$

Asi para todo x, se satisfaoe la siguiente desigualdad:

$$-\varepsilon - \alpha\phi_t(x) \leq f(t) - f(x) \leq \varepsilon + \alpha\phi_t(x)$$

Para escribir una desigualdad en las funciones. sea $f_o(x) = 1$. Entonces tenemos

$$-\varepsilon f_o - \alpha\phi_t \leq f(t)\, f_o - f \leq \varepsilon f_o + \alpha\, \phi_t$$

Por la linealidad y monotonicidad de L_n tenemos

$$-\varepsilon\,(L_n f_o)(t) - \alpha(L_n\phi_t)(t) \leq f(t)(L_n f_o)(t) - (L_n f)(t) \leq \varepsilon(L_n\, f_o)(t) + \alpha(L_n\, \phi_t)(t)$$

Esto produce $|f(t)(L_n f_o)(t) - (L_n f)(t)| \leq \varepsilon \, ||L_n f_o|| + \alpha(L_n \phi_t)(t)$.

Ya que $L_n f_o \to f_o$ y $(L_n \phi_t)(t) \to 0$, está claro que esta desigualdad finaliza esencialmente la prueba. Para ver exactamente cuan grande debe ser n, escriba

$$|f(t) - (L_n f)(t)| \leq |f(t) - f(t)(L_n f_o)(t)| + |f(t)(L_n f_o)(t) - (L_n f)(t)|$$

$$\leq |f(t)| \, |1 - (L_n f_o)(t)| + \varepsilon \, ||L_n f_o|| + \alpha(L_n \phi_t)(t)$$

$$\leq ||f|| \, ||f_o - L_n f_o|| + \varepsilon(1 + ||f_o - L_n f_o||) + \alpha(L_n \phi_t)(t)$$

Asi deberiamos seleccionar N de tal forma que siempre que $n \geq N$ resultará que $(||f|| + \varepsilon)||f_o - L_n f_o|| < \varepsilon$ y $\alpha(L_n \phi_t)(t) < \varepsilon$

PRUEBA DEL TEOREMA WEIERSTRASS

Vamos a probar el teorema para el intervalo $[0,1]$, dejando la extensión a un intervalo arbitrario para los problemas. Se mostrará que para cualquier $f \in C[0,1]$ los polinomios de Bernstein $B_n f$ (definidos en la pág. 69) convergen a f.

La linealidad y monotonicidad de B_n ya han sido mencionadas. Mediante el teorema de los operadores monótonos, será suficiente mostrar que $B_n f \to f$ para $f(x) = 1$, x y x^2.

Que $B_n 1 = 1$ resulta del teorema binomial:

$$(B_n 1)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1 - x)^{n-k} = [x + (1 - x)]^n = 1$$

Para la función $f(x) = x$ tenemos

$$(B_n f)(x) = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} x^k (1 - x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{kn!}{n(n-k)!\,k!} x^k(1-x)^{n-k}$$

$$= x \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}$$

$$= x \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k}(1-x)^{n-1-k}$$

$$= x\,[x + (1-x)]^{n-1} = x$$

Para la función $f(x) = x^2$ tenemos

$$(B_n f)(x) = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^2 \binom{n}{k} x^k(1-x)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{k}{n} \binom{n-1}{k-1} x^k(1-x)^{n-k}$$

$$= \frac{n-1}{n} \sum_{k=1}^{n} \frac{k-1}{n-1} \binom{n-1}{k-1} x^k(1-k)^{n-k} + \frac{1}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} x^k (1-x)^{n-k}$$

$$= \frac{n-1}{n} x^2 + \frac{1}{n} x \to x^2$$

Fue observado anteriormente que el procedimiento de interpolación de Lagrange, para un arreglo fijo de nodos, falla para proporcionar aproximaciones de precisión arbitrariamente altas para todas las funciones contínuas. Asi el Teorema Weierstrass no es una simple consecuencia de la fórmula de interpolación de Lagrange. Deberíamos esperar que esto sea verdadero en la totalidad para la fórmula de interpolación de Hermite puesto que en su aplicación usual esto involucra a las derivadas de la función. En 1930 Fejér hizo el descubrimiento sorprendente que el procedimiento de la interpolación de Hermite puede ser arreglado de tal forma para producir

una prueba del teorema Weierstrass.

En la fórmula de Hermite (pág 65) tomemos $y_i = f(x_i)$ y $y_i' = 0$. Al operador resultante lo llamaremos el operador Fejér - Hermite; este operador toma la forma

$$(5) \qquad (Lf)(x) = \sum_{i=1}^{n} f(x_i)\, A_i(x)$$

$$= \sum_{i=1}^{n} f(x_i) \left[1 - 2(x-x_i)\, \ell_i'(x_i) \right] \ell_i^2(x)$$

Para el propósito del teorema de Fejér será conveniente expresar el operador (5) en términos de la función $W(x) = \Pi(x - x_i)$. Tenemos, como en la sección previa,

$$\ell_i(x) = \frac{W(x)}{(x-x_i)W'(x_i)}$$

Si diferenciamos y luego usamos la regla de L'Hospital para evaluar la forma indeterminada obtenemos:

$$\ell_i'(x_i) = \frac{1}{2} W''(x_i)/W'(x_i)$$

Asi (5) toma la forma alternativa

$$(6) \qquad (Lf)(x) = \Sigma\, f(x_i) \left[1 - \frac{(x-x_i)W''(x_i)}{W'(x_i)} \right] \left[\frac{W(x)}{(x-x_i)W'(x_i)} \right]^2$$

TEOREMA [FEJER]

Denota L_n el operador Fejér - Hermite con nodos en los ceros del n-esimo polinomio Tchebycheff T_n.

Entonces $L_n f \to f$ para toda $f \in C[-1,1]$

PRUEBA. Korovkin 1959

Comenzamos estableciendo una fórmula para L_n en este caso especial:

$$(7) \qquad (L_n f)(x) = \frac{1}{n^2} T2_n(x) \sum_{i=1}^{n} f(x_i) \frac{1 - xx_i}{(x-x_i)^2}$$

donde los puntos x_i son los ceros de T_n, es decir,

$$x_i = \cos(2i - 1)\pi/2n$$

Una comparación de esta fórmula con la fórmula general (6) muestra que debemos establecer la igualdad.

$$(8) \qquad \frac{1}{n^2}(1 - x x_i) = \left[1 - (x-x_i) \frac{T''_n(x_i)}{T'_n(x_i)}\right] \frac{1}{[T'_n(x_i)]^2}$$

Puesto que $T_n(x) = \cos(n \cos^{-1}x)$, tenemos

$$T'_n(x) = n \operatorname{sen}(n \cos^{-1}x)(1 - x^2)^{-1/2} \qquad \text{y}$$

$$[T'_n(x_i)]^2 = n^2 \operatorname{sen}^2[(2i - 1)\pi/2](1 - x_i^2)^{-1} = n^2(1 - x_i^2)^{-1}.$$

Por otro lado empezando con la ecuación diferencial

$$(1-x^2) T''_n(x) - xT'_n(x) + n^2 T_n(x) = 0$$

(Problema 2, sec. 2), podemos colocar $x = x_i$ para deducir que

$(1-x_i^2) T''_n(x_i) = x_i T'_n(x_i)$ y $T''_n(x_i)/T'_n(x_i) = x_i/(1 - x_i^2)$.

Con estas sustituciones es fácil verificar la ecuación (8). Para completar la prueba usamos el teorema de los operadores monótonos. Que L_n es monótono es inmediato de (7) ya que $1 - xx_i \geq 0$. Mediante la parte de unicidad del teorema de interpolación de Hermite (sec. 2) resulta que $L_n 1 = 1$ De conformidad con el teorema de los operadores monótonos, solamente nece-

sitamos mostrar que $(L_n \phi_x)(x) \to 0$ uniformemente en x, donde $\phi_x(t) = (x-t)2$

Tenemos

$$(L_n \phi_x)(x) = \frac{1}{n^2} T^2{}_n(x) \sum_{i=1}^{n} (x-x_i)^2 \frac{1 - x x_i}{(x - x_i)^2}$$

La suma no puede exceder a 2n ya que $0 \leq 1 - x x_i \leq 2$.

Por lo tanto, la expresión completa converge a cero uniformemente - cuando $n \to \infty$

PROBLEMAS

1.- Pruebe que $\frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. De una prueba directa para la función

$f(x) = x^3$ que $B_n f \to f$

2.- Pruebe que si $f \in C[a,b]$ y $\phi(x) = f[a + x(b - a)]$, entonces $\phi \in C[0;1]$. Emplee este hecho para establecer el teorema Weierstrass para un intervalo cerrado arbitrario.

Si f es contínua en $(-\infty; +\infty)$ entonces los polinomios P_n existen - tal que para cada x, $P_n(x) \to f(x)$.

CAPITULO III

APROXIMACIÓN DE CUADRADOS MINIMOS Y TOPICOS RELACIONADOS

1. SISTEMAS ORTOGONALES DE POLINOMIOS

Dada una sucesión linealmente independiente de vectores $\{f_1, f_2, \ldots\}$ en cualquier espacio con producto interno, el proceso Gram-Schmidt (T.5 - de Apéndice) puede ser aplicado para generar un sistema ortonormal $\{g1, g2, \ldots\}$ con la propiedad que para cada n los subespacios generados por $\{f_1, \ldots, f_n\}$ y por $\{g1, \ldots, g_n\}$ son idénticos.

En esta sección investigamos los casos especiales que surgen cuando las funciones f_n son polinomios, en particular $f_n(x) = x^n$.

Comenzamos con un criterio simple para la independencia lineal de un conjunto de polinomios.

TEOREMA 1

Cualquier sucesión de polinomios $\{Q_o, Q1, \ldots\}$, en la cual (para cada n) Q_n es un polinomio de grado exacto n, es linealmente independiente. Un polinomio arbitrario de grado $\leq n$ es expresable unicamente como una combinación lineal de $Q_o, \ldots, Q_n$.

PRUEBA

La segunda afirmación implica la primera, porque si el polinomio cero puede ser escrito solamente en la forma $0 = 0Q_o + 0Q_1, +\ldots,$ entonces la sucesión $\{Q_o, Q1, \ldots\}$ es independiente. Si la segunda afirmación es-

falsa, sea n el primer índice para el cual falla. Ciertamente $n > 0$, ya - que toda constante es un múltiplo único de la constante no cero Q_0. Sea ahora P un polinomio arbitrario de grado $\leq n$, es decir

$$P(x) = \sum_{i=0}^{n} c_i x_i.$$

Esto debe ser expresado en la forma

$$P = \sum_{i=0}^{n} \lambda_i Q_i$$

El término x^n ocurre a la derecha solamente en Q_n y tiene, digamos, el coeficiente a_n. Entonces λ_n está determinada unicamente por el requerimiento de que $c_n x^n = \lambda_n a_n x^n$. Por la minimalidad de n y por el hecho que- $P - \lambda_n Q_n$ es de grado$< n$ resulta que $P - \lambda_n Q_n$ es unicamente expresable en la forma

$$\sum_{i=0}^{n-1} \lambda_i Q_i.$$

En la aplicación del proceso Gram-Schmidt a una sucesión {f1,f2...} cada miembro g_n del conjunto ortonormal es definido como una combinación lineal de f_n y todos precediendo a g_k. Una simplificación ocurre en el - caso de la sucesión {1, x, x2,...} de acuerdo con el siguiente resultado importante. En este teorema, el intervalo $[a,b]$ y la función peso w han sido prescritos, y el producto interno es definido como

$$\langle f,g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx.$$

TEOREMA 2

La sucesión de polinomios definida inductivamente de la manera sigui

ente es ortogonal:

$$Q_n = (x - a_n)\, Q_{n-1} - b_n\, Q_{n-2}$$

con $Q_o = 1$, $Q_1 = x - a_1$, $a_n = \langle xQ_{n-1}, Q_{n-1}\rangle / \langle Q_{n-1}, Q_{n-1}\rangle$,

y $b_n = \langle xQ_{n-1}, Q_{n-2}\rangle / \langle Q_{n-2}, Q_{n-2}\rangle$

PRUEBA

Observamos de las fórmulas que para cada n, Q_n es un polinomio mónico (es decir su coeficiente principal es la unidad) y no es por lo tanto cero. Por consiguiente los denominadores en las fórmulas para a_n y b_n no son cero.

Mostramos ahora por inducción en n que $\langle Q_n, Q_i\rangle = 0$ para $i < n$. Para n = 0 no hay nada que probar. Para n = 1 tenemos $\langle Q_1, Q_o\rangle = \langle xQ_o - a_1 Q_o, Q_o\rangle = \langle xQ_o, Q_o\rangle - \langle a_1 Q_o, Q_o\rangle = \langle xQ_o, Q_o\rangle - \langle xQ_o, Q_o\rangle = 0$. Ahora asumamos que nuestra afirmación es verdadera para n - 1. Entonces tenemos

$\langle Q_n, Q_{n-1}\rangle = \langle xQ_{n-1} - a_n Q_{n-1} - b_n Q_{n-2}, Q_{n-1}\rangle = \langle xQ_{n-1}, Q_{n-1}\rangle - a_n \langle Q_{n-1}, Q_{n-1}\rangle = 0$.

Similarmente $\langle Q_n, Q_{n-2}\rangle = \langle xQ_{n-1}, Q_{n-2}\rangle - a_n \langle Q_{n-1}, Q_{n-2}\rangle - b_n \langle Q_{n-2}, Q_{n-2}\rangle = 0$. Ahora si $i < n-2$, tenemos $\langle Q_n, Q_i\rangle = \langle xQ_{n-1}, Q_i\rangle - a_n \langle Q_{n-1}, Q_i\rangle - b_n \langle Q_{n-2}, Q_i\rangle = \langle Q_{n-1}, xQ_i\rangle = \langle Q_{n-1}, Q_{i+1} + a_{i+1} Q_i + b_{i+1} Q_{i-1}\rangle = 0$. Aquí hemos empleado la fórmula de recurrencia para obtener una expresión para $x\, Q_i$.

EJEMPLO

Si el proceso anterior es empleado en el intervalo $[-1,1]$ con

$$w(x) = 1,$$

los polinomios resultantes son llamados polinomios Legendre y son denotados por $X_0, X_1, \ldots$ Computemos los primeros de éstos, para ilustrar el uso de la relación de recurrencia.

i) $X_0 = 1$

ii) $a_1 = \langle xX_0, X_0\rangle / \langle X_0, X_0\rangle = 0$

iii) $X_1 = x$

iv) $a_2 = \langle xX_1, X_1\rangle / \langle X_1, X_1\rangle = 0$

v) $b_2 = \langle xX_1, X_0\rangle / \langle X_0, X_0\rangle = \frac{1}{3}$

vi) $X_2 = x^2 - \frac{1}{3}$

Los próximos dos polinomios Legendre son $X_3 = x^3 - \frac{3}{5}x$ y $X_4 = x^4 - \frac{6}{7}x^2 + \frac{3}{35}$.

COROLARIO

Si $f = \sum_{k=0}^{n} c_k Q_k$, entonces $f(x)$ puede ser evaluada con $2n-1$ multiplicaciones por medio de la fórmula de recurrencia $d_{n+2} = d_{n+1} = 0$, $d_k = c_k + (x - a_{k+1})d_{k+1} - b_{k+2} d_{k+2}$, $f(x) = d_0$.

PRUEBA

$$f(x) = \sum_{k=0}^{n} c_k Q_k(x) = \sum_{k=0}^{n} \left[d_k - (x-a_{k+1})d_{k+1} + b_{k+2}d_{k+2}\right] Q_k(x)$$

$$= d_0Q_0(x) + d_1\left[Q_1(x) - (x-a_1)Q_0(x)\right] + \sum_{k=2}^{n} d_k\left[Q_k(x) - (x-a_k)Q_{k-1}(x) + b_kQ_{k-2}(x)\right] = d_0.$$

Para contar el número de multiplicaciones requeridas observe que $d_n = c_n$ y

$d_{n-1} = c_{n-1} - (x - a_n)d_n$, mientras que la computación de $d_{n-2}, \ldots, d_o$, requiere cada una dos multiplicaciones.

Por lo tanto el total es 1 + 2(n-1) ó 2n-1.

Anteriormente cuando se consideraron los errores en la interpolación de Lagrange, probamos que el polinomio Tchebycheff $2^{-n+1}T_n$ tenía la norma Tchebycheff mínima en $[-1,1]$ entre todos los polinomios mónicos de grado n. Ahora vamos a ver que el mismo polinomio minimiza tambien la norma cuadrática

$$\left[\int_{-1}^{1} |f(x)|^2 \, (1 - x^{2)^{-}1/2} \, dx \right]^{1/2}$$

Además, este hecho emergerá como un caso especial de una afirmación más general como sigue:

TEOREMA 3

Los polinomios Q_n dados en el teorema 2 son los polinomios mónicos que hacen a la expresión $\langle \cdot, \cdot \rangle$ un mínimo.

PRUEBA

Ya que cada Q_n es mónico, un polinomio mónico arbitrario S es expresable en la forma $f = Q_n - a_{n-1} Q_{n-1} - \ldots - a_o Q_o$.

De conformidad con el teorema 7 del Apéndice, la cantidad $||f||^2 = \langle f,f \rangle$ será un mínimo si y sólo si los coeficientes a_k son los coeficientes de Fourier de Q_n con respecto al sistema ortogonal $\{ Q_o, \ldots, Q_{n-1} \}$ Asi $a_k = \langle Q_n, Q_k \rangle / \langle Q_k, Q_k \rangle = 0$, y $f = Q_n$ como se afirmó.

Una de las aplicaciones importantes de los polinomios ortogonales es al problema de integración numérica. Nos alejaremos brevemente de nuestro curso principal para dar una explicación de esta aplicación.

La integración numérica es el proceso de suponer el valor de una integral definida $\int_a^b f(x)\,dx$ de un número finito de valores o muestras de la función f. La definición de la integral como un límite de las sumas de Riemann, $\Sigma\, f(\xi_i)(x_{i+1} - x_i)$, sugiere que ésto debería ser posible. Miraremos los procesos lineales de este tipo. Ellos deben tomar la forma

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{n} A_k f(x_k) \tag{1}$$

Realmente podemos manejar el proceso de integración con una función-peso w(x) sin ninguna dificultad. Si los puntos de muestra x_k son fijos, los coeficientes A_k pueden ser determinados de acuerdo a diversos criterios de error, siendo el criterio usual que la fórmula será exacta para todos los polinomios de grado< n. Es posible arregalr esto integrando la fórmula de interpolación de Lagrange (pág. 60) con nodos x_k:

$$P(x) = \sum_{k=1}^{n} f(x_k)\ \ell_k(x)$$

$$\int_a^b f(x)\,w(x)\,dx \approx \int_a^b P(x)\,w(x)\,dx \tag{2}$$

$$= \sum_{k=1}^{n} f(x_k) \int_a^b \ell_k(x)\,w(x)\,dx$$

$$= \sum_{k=1}^{n} A_k \, f(x_k)$$

Si f es polinomio de grado < n, entonces P = f, y la fórmula de integración es exacta.

Nosotros debemos a Gauss el descubrimiento que mediante una colocación diestra de los nodos x_k la fórmula (2) puede ser hecha exacta para todos los polinomios de grado < 2n: Los x_k deben ser los ceros del polinomio ortogonal Q_n de grado n. Las fórmulas resultantes son llamadas "fórmulas (o cuadraturas) de integración Gaussianas".

TEOREMA 4

Sea exacta la fórmula de integración $\int_a^b f(x)\, w(x)\, dx \approx \sum_{k=1}^{n} A_k \, f(x_k)$ para todos los polinomios de grado < n. Será exacta para todos los polinomios de grado < 2n si y sólo si los nodos x_k son los ceros del polinomio Q_n definido en el teorema 2.

PRUEBA

Primero sean $x_1, \ldots, x_n$ los ceros de Q_n. Si f es un polinomio de grado < 2n, entonces por el algoritmo de la división podemos encontrar los polinomios P y R tal que $f = Q_n P + R$, siendo los grados de R y P < n. La fórmula de cuadratura es exacta para R, y Q_n es ortogonal a P. Por consiguiente

$$\int fw = \int Q_n \, Pw + \int Rw = \int Rw = \Sigma A_k \, R(x_k) = \Sigma A_k \, f(x_k).$$

Para el inverso, sea exacta la fórmula para todos los polinomios de

grado< 2n. Entonces es exacta para

$$f(x) = Q_k(x) \prod_{i=1}^{n} (x - x_i) \text{ si } k < n. \text{ Consecuentemente}$$

$$\int fw = \Sigma A_j f(x_j) = 0, \text{ y } \Pi(x - x_i) \text{ es ortogonal a ca-}$$

da Q_k para $k = 0, \ldots, n-1$. Asi $\Pi (x - x_i)$ es un múltiplo de Q_n.

Hemos dejado una brecha lógica en el desarrollo de la teoría de la - cuadratura Gaussiana. Si el dominio de la función f es el intervalo $[a,b]$ la fórmula de integración puede ser aplicada a f sólo si los puntos de - muestra x_k estan situados en $[a,b]$. Los teoremas siguientes llenan ésta - brecha.

TEOREMA 5.

Sea $\{Q_0, Q1, \ldots\}$ una sucesión de polinomios (sub-índices denotando grados) la cual es ortogonal con respecto a un producto interno

$$\langle f,g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx. \text{ Si f es cualquier fun-}$$

ción contínua en $[a,b]$ ortogonal a $Q_0, \ldots, Q_{n-1}$, entonces f debe cambiar- signo por lo menos n veces en (a;b) o desaparecer identicamente.

PRUEBA

Puesto que $f \perp Q_0$ y $Q_0 = 1$, $\int_a^b f(x)\, w(x)\, dx = 0$. Así, si $f \neq 0$, - entonces f debe cambiar signo por lo menos una vez en (a,b). Si f cambia signo menos que n veces, sean $r1 < r_2 < \ldots < r_k$ los puntos de (a,b) don- de f cambia signo. Entonces en cada intervalo $(a,r1), (r1,r2), \ldots, (r_k,b)$, f no cambia signo pero tiene signos opuestos en intervalos adyacentes. Ya que esta propiedad es compartida por el polinomio $P(x) = \prod_{i=1}^{k} (x - r_i)$,-

resulta que $\int_a^b f(x)\, P(x)\, w(x)\, dx \neq 0$. Pero esto es una contradicción ya - que P (siendo un polinomio de grado $k < n$) es una combinación lineal de - $Q_0, \ldots, Q_{n-1}$ y es por lo tanto ortogonal a f.

COROLARIO 1.

Sea $\{Q_0, \ldots\}$ un sistema ortogonal de polinomios (sub-índices denotando grados) en el intervalo $[a,b]$ con la función peso w. Entonces las raíces de Q_n son simples y estan situadas en (a,b).

PRUEBA

Esto resulta del teorema 5, porque Q_n es ortogonal a $Q_0, \ldots, Q_{n-1}$.

COROLARIO 2.

Sea $\{Q_0, \ldots\}$ un sistema ortogonal como en el corolario 1. Sea P - la aproximación mejor de cuadrados mínimos de la forma $P = \sum_{i=0}^{n} c_i Q_i$ a- alguna función contínua f. Entonces P interpola a f en por lo menos n + 1 puntos de (a,b).

PRUEBA.

Esto resulta del teorema 5 porque P - f es ortogonal a $Q_0, \ldots, Q_n$ y debe por lo tanto cambiar signos por lo menos n + 1 veces.

En la discusión anterior, el valor de n ha sido estático, y la notación no ha mostrado la dependencia de los coeficientes A_k y los nodos en n.

Para discutir la conducta de los errores cuando $n \to \infty$, escribamos la n-ésima fórmula Gaussiana en la forma

$$(3) \qquad \int_a^b f(x)\, w(x)\, dx \approx \sum_{k=1}^{n} A_{nk}\, f(x_{nk})$$

TEOREMA DE STIELTJES

Los errores en las fórmulas de cuadratura Gaussiana (3) convergen a cero (cuando $n \to \infty$) para toda función contínua f en $[\, a,b\,]$

PRUEBA

Comenzamos mostrando que $A_{nk} > 0$. Sea P el polinomio el cual resulta cuando el factor $x - x_{nk}$ es removido de Q_n. Ya que P^2 es de grado $< 2n$, la fórmula (3) sería exacta para él. Asi $0 < \int P^2 w = \sum_{i=1}^{n} A_{ni}\, P^2(x_{ni}) = A_{nk}\, P^2(x_{nk})$. Ya que x_{nk} es una raíz simple de Q_n, $P(x_{nk}) \neq 0$. Por consiguiente $A_{nk} > 0$.

También necesitamos el hecho que $\sum_{k=1}^{n} A_{nk}$ es independiente de n. Esto resulta de la observación que (3) es exacta para la función $f(x) = 1$, de tal modo que $\int_a^b w(x)\, dx = \sum_{k=1}^{n} A_{nk}$.

Para la prueba correcta, sea f una función contínua arbitraria en $[a,b\,]$, y sea ε un número positivo dado. Mediante el teorema de Weierstrass (Cáp.2, sec . 3), podemos determinar un polinomio P tal que

$$|f(x) - P(x)| < \varepsilon/c$$

en $[\, a,b\,]$, donde c denota la constante $2 \int_a^b w(x)\, dx$. Si el grado de P es

menor que 2n, la fórmula (3) es exacta para P. Consecuentemente mediante la desigualdad triangular

$$\left|\int fw - \Sigma A_{nk} f(x_{nk})\right| \leq \left|\int fw - \int Pw\right| + \left|\Sigma A_{nk} P(x_{nk}) - \Sigma A_{nk} f(x_{nk})\right|$$

$$\leq \int |f-P|\, w + \Sigma A_{nk} \left| P(x_{nk}) - f(x_{nk})\right|$$

$$\leq \frac{\varepsilon}{c}\left(\int w + \Sigma A_{nk}\right) = \varepsilon$$

El cálculo de los número A_i en la fórmula de cuadratura Gaussiana puede ser efectuada directamente de la definición

$$A_i = \int_a^b \ell_i(x)\, w(x) = \int_a^b \frac{Q_n(x)}{(x-x_i)Q'_n(x_i)}\, w(x)\, dx$$

Sin embargo, existe una alternativa más conveniente, la cual discutimos a continuación. Permítasenos definir una nueva sucesión de polinomios ϕ_o, $\phi_1, \ldots$, utilizando la misma relación de recurrencia como para Q_n,

$$\phi_n(x) = (x - a_n)\, \phi_{n-1}(x) - b_n\, \phi_{n-2}(x)$$

Pero con valores iniciales diferentes $\phi_o(x) = 0$ y $\phi_1(x) = b_1 = \int_a^b w(x)\,dx$.

Es claro que ϕ_n será un polinomio de grado n-1. La forma alternativa para A_i es entonces

$$A_i = \frac{\phi_n(x_i)}{Q'_n(x_i)}$$

y esto es una consecuencia directa del siguiente teorema, puesto que

$$Q_n(x_i) = 0.$$

TEOREMA

Con ϕ_n y Q_n como anteriormente, tenemos

$$\phi_n(x) = \int_a^b \frac{Q_n(t) - Q_n(x)}{t - x} w(t)\, dt.$$

PRUEBA

La prueba será por inducción. Nosotros comenzamos con el paso inductivo. Suponga que nuestra ecuación es verdadera para las integrales 0, 1, ..., n-1. Si $n \geq 2$, entonces nosotros tenemos por las relaciones de recurrencia

$$\int_a^b \frac{Q_n(t) - Q_n(x)}{t - x} w(t)\, dt$$

$$= \int_a^b \frac{(t-a_n)Q_{n-1}(t) - b_n Q_{n-2}(t) - (x-a_n)Q_{n-1}(x) + b_n Q_{n-2}(x)}{t - x} w(t)\, dt$$

$$= (x - a_n) \int_a^b \frac{Q_{n-1}(t) - Q_{n-1}(x)}{t - x} w(t)\, dt - b_n \int_a^b \frac{Q_{n-2}(t) - Q_{n-2}(x)}{t - x} w(t)\, dt$$

$$+ \int_a^b Q_{n-1}(t)\, w(t)\, dt$$

$$= (x - a_n)\, \phi_{n-1}(x) - b_n\, \phi_{n-2}(x)$$

$$= \phi_n(x)$$

Puesto que el argumento anterior requiere que $n \geq 2$, se hace necesario una verificación separada para $n = 0$ y $n = 1$. Para $n = 0$ nosotros observamos que ambos lados de la ecuación afirmada se reducen a cero. Para $n = 1$ el lado derecho se reduce a $\int_a^b w(t)\, dt$, y esto es $\phi 1(x)$ por definición.

PROBLEMAS

1.- La sucesión $\{1/\sqrt{2}, \cos x, \operatorname{sen} x, \cos 2x, \operatorname{sen} 2x, \ldots\}$ es ortonormal con respecto al producto interno

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, g(x)\, dx.$$

parte de la verificación es como sigue. De las identidades $\cos(A \pm B) = \cos A \cos B \mp \operatorname{sen} A \operatorname{sen} B$ obtenemos

$$\cos(A+B) + \cos(A-B) = 2 \cos A \cos B.$$

Por lo tanto

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \cos nx \cos mx\, dx =$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\cos(n+m)x + \cos(n-m)x\right] dx.$$

Si $n \neq m$, entonces ésto se convierte en

$$\frac{1}{2\pi} \left[\frac{\operatorname{sen}(x+m)x}{n+m} + \frac{\operatorname{sen}(n-m)x}{n-m} \right]_{-\pi}^{\pi} = 0$$

Si $n = m \geq 1$, esto llega a ser

$$\frac{1}{2\pi} \left[\frac{\operatorname{sen} 2nx}{2n} + x \right]_{-\pi}^{\pi} = 1$$

2.- La sucesión de polinomios de Tchebycheff

$$\left\{ \frac{T_o}{2}, T_1, T_2, \ldots \right\}$$

es ortonormal con respecto al producto interno

$$\langle f,g \rangle = \frac{2}{\pi} \int_{-1}^{1} f(x)\, g(x)\, \frac{dx}{\sqrt{1 - x^2}}$$

La verificación consiste en efectuar un cambio de variable $x = \cos \theta$

en la integral y usando el problema 1:

$$\frac{2}{\pi}\int_{-1}^{1} T_n(x)T_m(x)\,\frac{dx}{\sqrt{1-x^2}} = \frac{2}{\pi}\int_0^{\pi}\cos n\theta\cos m\theta\,d\theta = \begin{cases} 0 & (n\neq m) \\ 1 & (n=m\neq 0) \\ 2 & (n=m=0) \end{cases}$$

2. CONVERGENCIA DE EXPANSIONES ORTOGONALES

En el espacio lineal de las funciones contínuas en $[a,b]$ sea un producto interno definido por la ecuación

$$\langle f,g\rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx$$

y sea $\{\bar{Q}_0, \bar{Q}_1, \ldots\}$ el sistema ortonormal de los polinomios obtenidos mediante la aplicación del proceso Gram-Schmidt a $\{1, x, x^2, \ldots\}$. El teorema 7 del Apéndice nos dice que si una función dada f debe ser aproximada por una combinación lineal $\sum_{i=0}^{n} c_i \bar{Q}_i$ en la norma de los cuadrados mínimos, entonces los coeficientes c_i deben ser los coeficientes de Fourier de f: $c_i = \langle f,\bar{Q}_i\rangle$. Una característica extraordinaria de esta aproximación es que los coeficientes optimos c_i son independientes de n. Esto contrasta marcadamente con las aproximaciones Tchebycheff óptimas, en las cuales los coeficientes generalmente dependen de n.

Así a cada función continua f corresponde una expansión ortogonal formal, $\sum_{i=0}^{\infty} \langle f, \bar{Q}_i\rangle\, \bar{Q}_i$, la cual tiene la propiedad de que sus sumas parciales son las mejores aproximaciones a f en el sentido de los cuadros mínimos. No está claro por el momento si la serie converge en cualquier punto, o si representa f(x) en los puntos donde converge. Si f es un polinomio de grado n, entonces la serie converge uniformemente a f en virtud del hecho que

$$f = \sum_{k=0}^{n} \langle f, \bar{Q}_i\rangle\, \bar{Q}_i \quad \text{y} \quad \langle f, \bar{Q}_i\rangle = 0 \quad \text{para } i > n.$$

Para discutir las cuestiones de convergencia, definamos

$$S_n f = \sum_{i=0}^{n} \langle f, \bar{Q}_i\rangle\, \bar{Q}_i.$$

Además, denote $\mathfrak{J}_n f$ el polinomio de grado $\leq n$ el cual se aproxima mejor a f en el sentido Tchebycheff. La norma de cuadrados mínimos con peso w y la norma Tchebycheff serán distinguidas por los sub-índices w y T. Ahora-para las funciones contínuas f generalmente tenemos

$$||f - S_n f||_T \to 0$$

Este resultado es un corolario de un teorema más profundo del cual - no nos ocuparemos aquí. Para los otros tipos de convergencia prevalece, - sin embargo, una situación más favorable.

TEOREMA

Para todo $f \in C[a,b]$ tenemos

i) $||f - \mathfrak{J}_n f||_T \to 0$

ii) $||f - \mathfrak{J}_n f||_w \to 0$

iii) $||f - S_n f||_w \to 0$

PRUEBA

El teorema Weierstrass (Cap. 2 sec . 3) implica (i). El enunciado, - (ii) se sigue de (i) y de la siguiente desigualdad

$$\int_a^b |f(x) - (\mathfrak{J}_n f)(x)|^2 \, w(x) \, dx \leq ||f - \mathfrak{J}_n f||_T^2 \int_a^b w(x) \, dx$$

Finalmente el enunciado (iii) se sigue de (ii) y del hecho que $S_n f$ - es la mejor aproximación a f en la norma de los cuadrados mínimos, de tal forma que $||f - S_n f||_w \leq ||f - \mathfrak{J}_n f||_w$.

Las sumas parciales $S_n f$ de la expansión ortogonal $\sum \langle f, \overline{Q}_k \rangle \, \overline{Q}_k$ no

pueden solamente fallar para converger uniformemente para una función contínua particular f sino ellas pueden fallar para converger punto por punto. Es decir pueda que existan puntos ξ tal que $\lim_{n \to \infty} (S_n f)(\xi)$ no exista. Hay - aquí una cierta analogia con los polinomios de interpolación de Lagrange: usualmente imponiendo ciertas condiciones de uniformidad en f podemos probar que $||f - S_n f||_T \to 0$. El siguiente teorema ilustra esta observación,- aun cuando esta lejos de ser la mejor posible.

TEOREMA

Si f posee una segunda derivada contínua en $[-1,1]$, entonces la expansión de f en los polinomios Tchebycheff converge uniformemente a f.

PRUEBA

La expansión referida es $\frac{1}{2} A_o + \sum_{k=1}^{\infty} A_k T_k$, donde

$$A_k = \frac{2}{\pi} \int_{-1}^{1} f(x)\, T_k(x)\, \frac{dx}{\sqrt{1 - x^2}}$$

Haciendo el cambio de variable $x \to \cos \theta$, obtenemos

$$A_k = \frac{2}{\pi} \int_{o}^{\pi} g(\theta) \cos k\theta \, d\theta$$

donde $g(\theta) = f(\cos \theta)$. Integrando por partes dos veces en sucesión da:

$$A_k = \frac{-2}{\pi k^2} \int_{o}^{\pi} \cos k\theta \; g''(\theta) \, d\theta$$

De la hipótesis en f resulta entonces que A_k satisface una desigualdad de la forma $|A_k| \leq Mk^{-2}$. Asi la serie $\Sigma |A_k|$ converge y la serie-

Tchebycheff para f converge uniformemente por la prueba M de Weierstrass T.8 de Apéndice. Por otro teorema de Weierstrass, o por el teorema 9 del Apéndice, la función F a la cual la serie converge es contínua. Queda - probar que F = f. Dejemos que $S_n f$ denote la n-ésima suma parcial de la expansión de f. Entonces

$$||f - F||_w \leq ||f - S_n f||_w + ||S_n f - F||_w$$

El primer término de la derecha converge a cero, por el teorema anterior. El segundo término también llega a cero, porque $S_n f$ converge uniformemente a F. Por lo tanto $||f - F||_w = 0$ y $f = F$.

Como una ilustración de este teorema, considere a la función

$$f(x) = e^{\lambda x}.$$

Los coeficientes de la expansión $f = \frac{1}{2} A_o + \Sigma A_n T_n$ son dados por

$$A_n = \frac{2}{\pi} \int_0^{\pi} e^{\lambda \cos\theta} \cos n\theta \, d\theta$$

En términos de la función de Bessel, esta integral vuelve a ser

$$A_n = 2i^{-1} J_n(i\lambda) = 2I_n(\lambda).$$

Los primeros de estos coeficientes son como sigue cuando $\lambda = 1$.

$1/2\ A_o = 1.2660658778$	$A5 = 0.0005429263$
$A1 = 1.1303182080$	$A_6 = 0.0000449773$
$A_2 = 0.2714953395$	$A7 = 0.0000031984$
$A_3 = 0.0443368498$	$A8 = 0.0000001992$
$A4 = 0.0054742404$	$A_9 = 0.0000000110$

Para dar una discusión general de expansiones en polinomios ortogo

nales, es conveniente interpretar el operador S_n, definido por la ecuación

$$S_n f = \sum_{i=0}^{n} \langle f, \overline{Q}_i \rangle \, \overline{Q}_i$$

como un operador integral. En verdad nosotros tenemos

$$(1) \qquad (S_n f)(x) = \sum_{i=0}^{n} \int_a^b f(t) \, \overline{Q}_i(t) \, w(t) \, dt \, \overline{Q}_i(x)$$

$$= \int_a^b f(t) \sum_{i=0}^{n} \overline{Q}_i(t) \, \overline{Q}_i(x) \, w(t) \, dt$$

$$= \int_a^b f(t) \, K_n(t,x) \, w(t) \, dt$$

donde tenemos el conjunto, $K_n(t,x) = \sum_{i=0}^{n} \overline{Q}_i(t) \, \overline{Q}_i(x)$. Esta función se le llama el núcleo del sistema ortonormal. Lo que acabamos de escribir no está para nada limitado a los sistemas ortonormales de polinomios, sino que para los teoremas posteriores supondremos que los $\overline{Q}_n$ se obtienen aplicando el proceso de Gram-Schmidt a $\{1, x, x2, \ldots\}$. Los polinomios $Q_n = \lambda_n \overline{Q}_n$ son los polinomios mónicos discutidos en la sección anterior. En otras palabras, λ_n^{-1} es el coeficiente de x^n en $\overline{Q}_n$.

LEMA (Identidad de Christoffel - Darboux)

El núcleo toma la forma

$$\sum_{i=0}^{n} \overline{Q}_i(x) \, \overline{Q}_i(t) = \lambda_{n+1} \, \lambda_n^{-1} \, \frac{\overline{Q}_{n+1}(x)\overline{Q}_n(t) - \overline{Q}_n(x)\overline{Q}_{n+1}(t)}{x - t}$$

PRUEBA

De la relación de recurrencia de tres términos (Teorema 2 sec. 1), tenemos las ecuaciones

$$Q_{n+1}(x)\, Q_n(t) = (x - a_{n+1})\, Q_n(x)\, Q_n(t) - b_{n+1} Q_{n-1}(x)\, Q_n(t)$$

$$Q_{n+1}(t)\, Q_n(x) = (t - a_{n+1})\, Q_n(t)\, Q_n(x) - b_{n+1} Q_{n-1}(t)\, Q_n(x)$$

Si nosotros restamos la segunda de estas ecuaciones de la primera, obtenemos

$$(2) \qquad Q_{n+1}(x)\, Q_n(t) - Q_{n+1}(t)\, Q_n(x)$$

$$= (x-t)\, Q_n(t)\, Q_n(x) + b_{n+1}\left[Q_n(x)\, Q_{n-1}(t) - Q_n(t)\, Q_{n-1}(x)\right]$$

Note que $\lambda_n = \sqrt{\langle Q_n, Q_n \rangle}$ puesto que $Q_n = \lambda_n \overline{Q}_n$. De la fórmula de recurrencia escrita en la forma $b_{n+1} Q_{n-1} = x\, Q_n - a_{n+1} Q_n - Q_{n+1}$ nosotros obtenemos, tomando producto interno con Q_{n-1},

$$b_{n+1}\, \lambda_{n-1}^2 = \langle x\, Q_n, Q_{n-1} \rangle = \langle Q_n, x\, Q_{n-1} \rangle$$

$$= \left\langle Q_n, Q_n + a_n Q_{n-1} + b_n Q_{n-2} \right\rangle = \lambda_n^2 \qquad (Q_{-1} = 0)$$

Por lo tanto si la ecuación (2) se divide por λ_n^2 el resultado es

$$(3) \qquad \lambda_n^{-2}\left[Q_{n+1}(x)\, Q_n(t) - Q_{n+1}(t)\, Q_n(x)\right]$$

$$= (x - t)\, \overline{Q}_n(x)\, \overline{Q}_n(t) + \lambda_{n-1}^{-2}\left[Q_n(x)\, Q_{n-1}(t) - Q_n(t)\, Q_{n-1}(x)\right]$$

Ahora reaplicamos la fórmula de recurrencia (3) a fin de simplificar el último término. Eventualmente nosotros obtenemos

$$\lambda_n^{-2}\left[Q_{n+1}(x)\, Q_n(t) - Q_{n+1}(t)\, Q_n(x) \right]$$

$$= (x - t) \sum_{i=1}^{n} \overline{Q}_i(x)\, \overline{Q}_i(t) + \lambda_o^{-2}\left[Q_1(x)\, Q_o(t) - Q_1(t)\, Q_o(x) \right]$$

$$= (x - t) \sum_{i=0}^{n} \overline{Q}_i(x)\, \overline{Q}_i(t)$$

La prueba se completa escribiendo $\lambda_n \overline{Q}_n$ por Q_n, etc.

DESIGUALDAD DE BESSEL

En un espacio con producto interno, sea $\{g_n\}$ cualquier sucesión ortonormal, y sea f cualquier elemento. Entonces $\Sigma \langle f, g_n \rangle^2 \leq \langle f, f \rangle$. En particular, los coeficientes de Fourier de f convergen a cero.

PRUEBA

Poner $F_n = \sum_{i=0}^{n} \langle f, g_i \rangle\, g_i$. Escribiremos $f \perp g$ cuando $\langle f, g \rangle = 0$ y decimos que "f es ortogonal a g", así $f - F_n \perp F_n$. En realidad, puesto que f es una combinación lineal de $g_0, \ldots, g_n$, necesitamos unicamente observar (para $j \leq n$) que $\langle f - F_n, g_j \rangle = \langle f, g_j \rangle - \sum_{i=0}^{n} \langle f, g_i \rangle \langle g_i, g_j \rangle = 0$. Consecuentemente por la ley de Pitágoras (ver T.10 de Apéndice), tenemos

$$||f||^2 = ||F_n||^2 + ||f - F_n||^2 \geq ||F_n||^2 = \sum_{i=0}^{n} \langle f, g_i \rangle^2.$$

Puesto que esto es verdadero para todo n, resulta cierto en el límite.

TEOREMA

Si los valores de los polinomios $\overline{Q}_n$ quedan acotados en un cierto punto de $x_o \in [a,b]$, si f es contínua, y si f satisface en x_o la condición de Lipschitz $|f(x_o) - f(x)| \leq \alpha |x_o - x|$, entonces en ese punto $f(x_o) = \Sigma \langle f, \overline{Q}_n \rangle\, \overline{Q}_n(x_o)$.

PRUEBA

Nosotros observamos primero que los números $\lambda_n \lambda_{n-1}^{-1}$ son acotados. En realidad,

$$\lambda_n^2 = \langle Q_n, Q_n \rangle = \langle Q_n, x\,Q_{n-1} - a_n Q_{n-1} - b_n Q_{n-2} \rangle = \langle Q_n, x\,Q_{n-1} \rangle$$

$$\leq \int_a^b |x|\ |Q_n(x)|\ |Q_{n-1}(x)|\ w(x)\,dx \leq c\ \langle |Q_n|, |Q_{n-1}| \rangle$$

$$\leq c\,\|Q_n\|\ \|Q_{n-1}\| = c\lambda_n \lambda_{n-1}$$

Entonces $\lambda_n \lambda_{n-1}^{-1} \leq c \equiv \max_{a \leq x \leq b} |x|$. Ahora de la observación que $(S_n 1)(t) = 1$, resulta que el error en x_o es

$$\varepsilon_n \equiv f(x_o) - (S_n f)(x_o) = f(x_o)(S_n 1)(x_o) - (S_n f)(x_o)$$

De la forma integral de S_n [ecuación (1)] tenemos

$$\varepsilon_n = \int_a^b \left[f(x_o) - f(x) \right] \sum_{i=0}^{n} \overline{Q}_i(x_o)\, \overline{Q}_i(x)\, w(x)\,dx$$

Por la identidad de Christoffel - Darboux esto se vuelve

$$\varepsilon_n = \lambda_{n+1} \lambda_n^{-1} \int_a^b \frac{f(x_o) - f(x)}{x_o - x} \left[\overline{Q}_{n+1}(x_o)\, \overline{Q}_n(x) - \overline{Q}_n(x_o)\, \overline{Q}_{n+1}(x) \right] w(x)\,dx$$

$$= \lambda_{n+1} \lambda_n^{-1} \left[\langle h, \overline{Q}_n \rangle\, \overline{Q}_{n+1}(x_o) - \langle h, \overline{Q}_{n+1} \rangle\, \overline{Q}_n(x_o) \right]$$

donde $h(x) = [f(x_o) - f(x)]/(x_o - x)$. Mediante la condición de Lipschitz en f, $|h(x)| \leq \alpha$. También tenemos $\lambda_{n+1} \lambda_n^{-1} \leq c$.

Finalmente nosotros observamos que de la desigualdad de Bessel los coeficientes de Fourier de h, $\langle h, \overline{Q}_n \rangle$, tienden a cero. Puesto que los números $\overline{Q}_n(x_o)$ permanecen acotados, $\varepsilon_n \to 0$.

Concluiremos esta sección con dos teoremas célebres relacionados - con la convergencia de las series ordinarias de Fourier. De nuevo ciertos operadores integrales juegan un papel. Denotemos la n-ésima suma parcial- de la serie de Fourier de f por el símbolo $S_n f$:

$$(S_n f)(x) = \frac{a_o}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \operatorname{sen} kx)$$

donde

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt \, dt$$

y

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \operatorname{sen} kt \, dt$$

De acuerdo con nuestras observaciones primarias acerca de los sistemas ortonormales generales este operador S_n puede colocarse en forma - integral. El núcleo que ocurre acá se conoce como el núcleo Dirichlet.

LEMA

El operador S_n de la serie de Fourier tiene la siguiente forma integral para las funciones f contínuas y períodicas, con período 2π,

$$(4) \qquad (S_n f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t + x) \, \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} (\frac{t}{2})} \, dt$$

PRUEBA

En la definición de S_n, sustituya los integrales mediante los cuales a_k y b_k son definidos. El resultado es

$$(S_n f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^{n} (\cos kx \cos kt + \operatorname{sen} kx \operatorname{sen} kt) \right] dt$$

La aplicación de la identidad trigonométrica $\cos(A - B) = \cos A \cos B$

+ sen A sen B produce

$$(S_n f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^{n} \cos k(x - t) \right] dt$$

Ya que el integral es periódico nosotros podemos integrar en cambio sobre el intervalo $[-\pi - x, \pi - x]$. Entonces el cambio de la variable $t \to x-t$ nos da

$$(S_n f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t-x)(\frac{1}{2} + \sum_{k=1}^{n} \cos kt)\, dt$$

La prueba continúa ahora con una verificación de la ecuación

$$(\frac{1}{2} + \sum_{k=1}^{n} \cos kt)\, 2 \operatorname{sen} \frac{t}{2} = \operatorname{sen}(n + \frac{1}{2})t \tag{5}$$

Esto es verdadero porque con la ayuda de la identidad 2 cos A sen B = sen (A + B) - sen (A - B), el miembro de la izquierda de (5) llega a ser

$$\operatorname{sen} \frac{t}{2} + \sum_{k=1}^{n} \left[\operatorname{sen}(k + \frac{1}{2})t - \operatorname{sen}(k - \frac{1}{2})t = \operatorname{sen}(n + \frac{1}{2})t \right]$$

La ecuación (5) establece la forma correcta del núcleo en (4) en todos los puntos donde $\operatorname{sen}(\frac{t}{2}) \neq 0$. En los puntos donde $\operatorname{sen}(\frac{t}{2}) = 0$, un valor que limita debe ser tomado, siendo la existencia del límite una consecuencia de la continuidad de las funciones en (5).

TEOREMA

Si f es 2π - periódica contínua en cualquier parte, y diferenciable en x, entonces su serie de Fourier converge en x a f(x).

PRUEBA

Puesto que $S_n 1 = 1$, tenemos del lema,

$$\varepsilon_n \equiv (S_n f)(x) - f(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} \left[f(t - x) - f(x) \right] \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} \frac{1}{2} t} \, dt$$

Sea x fijo, y coloque $g(t) = \left[f(t + x) - f(x) \right] / (2 \operatorname{sen} \frac{1}{2} t)$. Esta función es contínua excepto en $t = 0$, donde no está definida. Pero

$$\lim_{t \to 0} g(t) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} \cdot \frac{t}{2 \operatorname{sen} \frac{1}{2} t} = f'(x)$$

Así que simplemente definiendo $g(0) = f'(x)$ hacemos contínua a g. Puesto que

$$\frac{1}{\pi} \int_{-\pi}^{\pi} g(t) \operatorname{sen} (n + \tfrac{1}{2})t \, dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} g(t) \cos \tfrac{1}{2}t \operatorname{sen} nt \, dt + \frac{1}{\pi} \int_{-\pi}^{\pi} g(t) \operatorname{sen} \tfrac{1}{2}t \cos nt \, dt$$

tenemos aquí los coeficientes de Fourier de las funciones $g(t) \cos \frac{1}{2} t$ y $g(t) \operatorname{sen} \frac{1}{2} t$. Estos coeficientes van hacia cero cuando $n \to \infty$ por la desigualdad de Bessel. Por lo tanto $\varepsilon_n \to 0$.

Un teorema similar al anterior fue probado allá por el año 1829 por Dirichlet. Su resultado fue que la serie Fourier de una función que es contínua y monótona en secciones converge en cada punto x a f(x) a condición de que en un punto de discontinuidad, f(x) fuera definida como

$$\frac{1}{2} \left[f(x+0) + f(x - 0) \right]$$

Por casi 50 años después de esto se congeturó que toda función contínua estaba representada por su serie de Fourier. Sin embargo, en 1876 Du Bois-Reymond tuvo éxito en construir un ejemplo de una función contínua cuya-

serie de Fourier divergía en un punto. Todavía es una pregunta abierta si una función contínua puede existir para lo cual la serie de Fourier diverge en todos los puntos de un intervalo. A pesar de estos resultados negativos, una transformación muy simple de la serie de Fourier puede ser utilizada para suministrar aproximaciones "uniformes" de precisión arbitraria para todas las funciones contínuas. Este descubrimiento por Fejér en 1900 puede ser enunciado facilmente: Los primeros medios Cesáro de la serie de Fourier de una función contínua convergen uniformemente a la función. Los medios Cesáro de una sucesión (g_n) son los promedios $G_n = \frac{1}{n} \sum_{k=1}^{n} g_k$.

Los medios Cesáro de una serie son entonces los promedios de sus sumas parciales. En el caso de la serie de Fourier, si retenemos el símbolo $S_n f$ para denotar las sumas parciales, entonces los medios Cesáro son las funciones

$$G_n f = \frac{1}{n} \left[S_o f + S_1 f + \ldots + S_{n-1} f \right]$$

Podemos poner este operador en la forma de un operador integral como sigue, el núcleo para el cual ocurre conocido como el núcleo Fejér.

LEMA

El operador Fejér G_n tiene la forma alternativa

$$(G_n f)(x) = \frac{1}{2n\pi} \int_{-\pi}^{\pi} f(t+x) \left(\frac{\operatorname{sen} \frac{1}{2} nt}{\operatorname{sen} \frac{1}{2} t} \right)^2 dt$$

PRUEBA

Usando la forma integral de los operadores S_n (pág. 101), tenemos

$$(G_n f)(x) = \frac{1}{n} \sum_{k=o}^{n-1} \frac{1}{\pi} \int_{-\pi}^{\pi} f(t + x) \frac{\text{Sen}(k + \frac{1}{2})t}{2 \text{ sen } \frac{1}{2} t} dt$$

$$= \frac{1}{2n\pi} \int_{-\pi}^{\pi} f(t + x) \sum_{k=o}^{n-1} \frac{\text{sen}(k + \frac{1}{2})t}{\text{sen } \frac{1}{2} t} dt$$

Para completar la prueba debemos mostrar que

$$\sum_{k=o}^{n-1} \frac{\text{sen}(k + \frac{1}{2}) t}{\text{sen } \frac{1}{2} t} = \left(\frac{\text{sen } \frac{1}{2} nt}{\text{sen } \frac{1}{2} t}\right)^2$$

Será suficiente establecer que

$$\sum_{k=o}^{n-1} \text{sen}(k + \tfrac{1}{2})t \text{ sen } \tfrac{1}{2} t = (\text{sen } \tfrac{1}{2} nt)^2$$

Usando la identidad 2 sen A sen B = cos (A - B) - cos (A + B), podemos escribir el miembro izquierdo de esta ecuación como

$$\frac{1}{2} \sum_{k=o}^{n-1} \left[\cos kt - \cos(k+1)t\right] - \frac{1}{2}(1 - \cos nt)$$

De la fórmula del"semi-ángulo", el último es $(\text{sen } \frac{1}{2} nt)2$.

Para probar el teorema Fejér necesitamos el análogo trigonométrico del teorema operador-monótono (Cap. 2 sec. 3). Nosotros lo establecemos sin prueba.

TEOREMA DE KOROVKIN

Dejemos que $\{L_n\}$ denote una sucesión de operadores lineales monótonos en $C_{2\pi}$. Para que $L_n f \to f$ (uniformemente) para toda $f \in C_{2\pi}$, es necesario y suficiente que tal convergencia ocurra para f = 1, cos y sen.

TEOREMA DE FEJER

Los primeros medios Cesáro de la serie de Fourier de una función continua y 2π-periódica convergen uniformemente a la función.

PRUEBA

Observamos del lema anterior que los operadores Fejér G_n son operadores monótonos; es decir, si $f \geq g$, entonces $G_n f \geq G_n g$. Por el teorema de Korovkin, podemos completar la prueba verificando que $G_n f \to f$ cuando $f = 1$, cos ó sen. Calculamos entonces,

$$G_n 1 = \frac{1}{n}(1 + \ldots + 1) = 1 \to 1$$

$$(G_n \cos)(x) = \frac{1}{n}(0 + \cos x + \ldots + \cos x) = \frac{n-1}{n} \cos x \to \cos x$$

$$(G_n \operatorname{sen})(x) = \frac{1}{n}(0 + \operatorname{sen} x + \ldots + \operatorname{sen} x) = \frac{n-1}{n} \operatorname{sen} x \to \operatorname{sen} x$$

3. LA APROXIMACION MEDIANTE SERIES DE POLINOMIOS DE TCHEBYCHEFF

La posibilidad de representar las funciones de la forma $f = \sum_{k=0}^{\infty} a_k T_k$ se consideró brevemente en la sección precedente, junto con los ejemplos de las expansiones en otros sistemas ortogonales. Se verá en breve que los polinomios de Tchebycheff tienen ventajas decisivas sobre otros sistemas ortogonales si es nuestro propósito proveer de buenas aproximaciones en la norma uniforme. En realidad para muchas funciones f, una expansión Tchebycheff truncada $S_n f = \sum_{k=0}^{n} a_k T_k$ es muy cercano a una mejor aproximación polinomial en la norma Tchebycheff. Por supuesto, $S_n f$ es la mejor aproximación de f en la norma de los cuadrados mínimos:

$$||f - S_n f||_w^2 = \int_{-1}^{1} |f(x) - (S_n f)(x)|^2 \, (1 - x^2)1/2 \, dx$$

Para las funciones que son altamente "regulares" (poseen muchas derivadas) la norma Tchebycheff,

$$||f - S_n f||_T = \max_{-1 \leq x \leq 1} |f(x) - (S_n f)(x)|$$

esta generalmente dentro de un pequeño porcentaje de su mínimo absoluto. Cuando tomamos en cuenta el hecho de que un polinomio de mejor aproximación es a menudo difícil de obtener, mientras que $S_n f$ es relativamente fácil de obtener, reconocemos que las expansiones Tchebycheff son un instrumento importante en las aproximaciones uniformes. Realmente, aún si se asume que f es solamente contínua (y si $n < 400$), nunca podemos obtener más de una cifra decimal extra de exactitud al pasar de $S_n f$ a el polinomio de mejor aproximación (Problema 1, sec. 5)

Para hacer comparaciones del tipo que se acaba de mencionar, empleamos nuevamente la notación $E_n(f)$ para la distancia (en la norma uniforme)-

de f al subespacio de polinomios que tienen grado $\leq n$. Así

$$E_n(f) = \min_{c_0,\ldots,c_n} \max_{-1\leq x\leq 1} \left| f(x) - \sum_{i=0}^{n} c_i x^i \right|$$

Con esta notación, el teorema de Weierstrass establece simplemente que $E_n(f) \to 0$ para toda $f \in C[-1,1]$. Después estableceremos algunos teoremas de Jackson los cuales afirman que $E_n(f)$ converge a cero todo lo más rápido posible cuando f es uniforme. En el momento que empleamos las expansiones Tchebycheff para establecer un teorema de "aletargamiento": $E_n(f)$ puede converger a cero muy despaciosamente para algunas funciones contínuas.

TEOREMA 1

Si $\{\varepsilon_n\}$ es cualquier sucesión convergiendo hacia abajo a cero, entonces, existe un elemento $f \in C[-1,1]$ tal que $E_n(f) \geq \varepsilon_n$ para todo n.

PRUEBA

Defina $a_k = \varepsilon_{k-1} - \varepsilon_k$ y $f = \sum_{k=0}^{\infty} a_k T_{3^k}$. Por hipótesis, $a_k \geq 0$. - Puesto que $||T_n|| = 1$, la serie para f es mayorizada por la serie Σa_k y - consecuentemente converge uniformemente por la prueba M de Weierstrass (T. 8 de Apéndice) Resulta que $f \in C[-1,1]$. Mostramos ahora que la mejor - aproximación de grado $\leq 3^n$ a f es simplemente la suma parcial

$$P = \sum_{k=0}^{n} a_k T_{3^k}$$

Será suficiente mostrar que la función de error $r = f - P = \sum_{k=n+1}^{\infty} a_k T_{3^k}$

obtiene su desviación máxima desde cero con signos que alternan en al menos $3^n + 2$ puntos del intervalo $[-1,1]$. Considere los puntos

$$x_i = \cos(i\pi/3^{n+1}) \quad \text{con } i = 0, \ldots, 3^{n+1}.$$

Si $k \geq n+1$, entonces $T_{3^k}(x_i) = \cos(3^k\, i\pi/3^{n+1}) = (-1)^i$. Asi $r(x_i) = (-1)^i \sum_{k=n+1}^{\infty} a_k = (-1)^i\, \varepsilon_n$. Puesto que

$$\varepsilon_n = |r(x_i)| \leq ||r|| \leq \sum_{k=n+1}^{\infty} |a_k| = \varepsilon_n ,$$

hemos probado en realidad que r alterna por lo menos $3^{n+1} + 1$ veces entre los valores $\pm\ \varepsilon_n$. Por consiguiente, $E_n(f) \geq E_{3^n}(f) = \varepsilon_n$.

Debería observarse que la prueba permanece válida si la sucesión $\{3^n\}$ es reemplazada por cualquier sucesión creciente de números enteros impares.

Ya que se conoce un teorema más general del mismo tipo, lo estableceremos aquí pero referimos al lector a [Timan 1960] ó [Golomb, 1960] para su prueba.

TEOREMA 2

Sea $\{g_1, g_2, \ldots\}$ una sucesión linealmente independiente en un espacio de Banach B. Si $\{\varepsilon_n\}$ es cualquier sucesión de números que convergen hacia abajo a cero, entonces existe una $f \in B$ tal que, para toda n,

$$\varepsilon_n = \inf_c \left|\left| f - \sum_{i=1}^{n} c_i g_i \right|\right|$$

En la prueba del teorema 1, se excibió una clase de funciones $f \in C[-1,1]$ que tienen la propiedad $P_n f = S_n f$

para toda n, $P_n f$ y $S_n f$ siendo respectivamente las mejores aproximaciones en la norma uniforme y en la norma de los cuadrados mínimos con peso - $(1 - x^2)^{-1/2}$. Sería un estado muy deseable de asuntos si muchas funciones comunes que requieren aproximación poseyeran esta propiedad. Desafortunadamente este no es el caso. Además, las funciones consideradas en el teorema 1 pueden ser altamente irregulares. El ejemplo clásico de ésto - (en términos de las series trigonométricas) es la función de Weierstrass,

$$(1) \qquad f(x) = \sum_{k=0}^{\infty} a^k \cos b^k x$$

en la cual $0 < a < 1$, y b es un número entero impar mayor que a^{-1}. Es claro que esta función es contínua en todas partes. Pero no es diferenciable en ningún punto sea como sea. Una prueba de esto (en una forma ligeramente debilitada) es incluida aquí a causa de su interés general.

TEOREMA 3

Si $0 < a < 1$ y si b es un número entero impar mayor que $6a^{-1}$, entonces la función Weierstrass (1) no es diferenciable en ningún lugar.

PRUEBA (Titchmarsh)

Sea x un punto arbitrario. Se probará el teorema mostrando como hacer que $h \to 0$ en tal forma que $h^{-1} |f(x+h) - f(x)| \to \infty$

Primero escriba, con abreviaciones obvias,

$$\left|\frac{f(x+h) - f(x)}{h}\right| = \left| \sum_{k=0}^{\infty} h^{-1} a^k [\cos b^k(x+h) - \cos b^k x] \right|$$

$$= \left| \sum_{k<n} A_k + \sum_{k\geq n} A_k \right|$$

$$\geq \left| \sum_{k \geq n} A_k \right| - \sum_{k<n} |A_k|$$

Con n fijo, seleccione un número entero ν tal que el número $h = -x + \nu\pi b^{-n}$ está situado en el intervalo

$$\left[\frac{1}{2}\pi b^n, \ \frac{3}{2}\pi b^{-n}\right).$$

Para $k \geq n$ tenemos $\cos b^k(x+h) = \cos b^{k-n} b^n(x+h) = \cos b^{k-n}\nu\pi = (-1)^\nu$ Asi los términos A_k, para $k \geq n$, todos tienen el mismo signo, a saber $(-1)^\nu$. Por consiguiente

$$\left| \sum_{k \geq n} A_k \right| = \sum_{k \geq n} |A_k| \geq |A_n|$$

ya que $-\cos b^n x = -\cos(\nu\pi - b^n h) = -(-1)^\nu \cos b^n h$ y $b^n h \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right)$ resulta que $-\cos b^n x$ tiene el signo $(-1)^\nu$. Por consiguiente

$$|A_n| = h^{-1}a^n \left|\cos b^n(x+h) - \cos b^n x\right|$$

$$\geq h^{-1}a^n \geq \frac{2}{3\pi} a^n b^n$$

Por otra parte, para $k < n$ podemos usar el teorema del valor medio para escribir

$$\sum_{k<n} |A_k| = \sum_{k<n} h^{-1}a^k \left|hb^k \operatorname{sen} b^k \varepsilon_k\right| \leq \sum_{k<n} a^k b^k \leq \frac{a^n b^n}{ab-1}$$

Combinando las desigualdades anteriores produce

$$\left| \sum_{k=0}^{\infty} A_k \right| \geq \left(\frac{2}{3\pi} - \frac{1}{ab-1}\right) a^n b^n$$

Está claro que para ab grandes el término entre paréntesis es positivo, y la expresión completa llegará a ser infinita cuando $n \to \infty$. Específicamente, uno puede chequiar que $b > 6a^{-1}$ es una condición suficiente.

Las series de Tchebycheff que hemos considerado hasta ahora en esta sección han tenido un comportamiento de convergencia muy simple: la serie $f = \Sigma\, a_k\, T_k$ ha convergido porque $\Sigma\, |a_k|$ convergió, y la prueba M Weierstrass fue aplicable. Los criterios de convergencia más sofisticados, involucrando a la función f, son disponibles. Es adecuado establecer uno de estos aqui, aún cuando por razones de economía la prueba es remitida a la sección 5.

Como se dijo antes, w denota al módulo de continuidad de la función f: $w(\delta) = \max_{|x-y| \leq \delta} |f(x) - f(y)|$.

TEOREMA DE DINI-LIPSCHITZ

Si la función f satisface la condición de Dini-Lipschitz en $[-1,1]$, $\lim_{\delta \to 0} w(\delta) \log \delta = 0$ entonces el desarrollo en polinomios Tchebycheff converge uniformemente en él.

Nosotros concluimos esta sección con algunos teoremas, los cuales nos capacitan para estimar $E_n(f)$ de una expansión formal de f en los polinomios ortogonales. Suponemos por lo tanto que tenemos una sucesión de polinomios $\{Q_0, Q_1, \ldots\}$ (los sub-índices indican el grado) la cual es ortogonal con respecto al producto interno

$$\langle f,g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx$$

Cualquier función $f \in C[a,b]$ entonces posee una expansión formal

$$f \sim \sum_{k=0}^{\infty} c_k\, Q_k$$

donde los coeficientes estan dados por $c_k = \langle f,Q_k\rangle / \langle Q_k, Q_k\rangle$. La serie puede o no converger a f.

TEOREMA 4

Sea f formalmente expandida en una serie de polinomios ortogonales, $f \sim \Sigma\, c_k\, Q_k$. Entonces

i) $E_n(f) \geq \max \{a_{n+1}\,|c_{n+1}|, a_{n+2}\,|c_{n+2}|, \ldots\}$

ii) $E_n(f) \geq \sqrt{\beta_{n+1}\, c_{n+1}^2 + \beta_{n+2}\, c_{n+2}^2 + \ldots}$

iii) $E_n(f) \leq \gamma_{n+1}\,|c_{n+1}| + \gamma_{n+2}\,|c_{n+2}| + \ldots$

donde

$$a_k = \int_a^b Q_k^2\, w \,/\, \int_a^b |Q_k|\, w,$$

$$\beta_k = \int_a^b Q_k^2\, w \,/\, \int_a^b w,$$

y $$\gamma_k = \max_{a<x<b} |\,Q_k(x)|$$

PRUEBA

Sea P el polinomio de grado $\leq$ n el cual se aproxima en mejor forma a f en la norma uniforme. Si k > n, entonces las relaciones de ortogonalidad implican

$$|c_k| \int_a^b Q_k^2 \, w = \left| \int_a^b f \, Q_k \, w \right|$$

$$= \left| \int_a^b (f - P) Q_k \, w \right|$$

$$\leq \int_a^b |f - P| \, |Q_k| \, w$$

$$\leq E_n(f) \int_a^b |Q_k| \, w$$

Esto prueba (i). Para probar (ii) sea $S_n = \sum_{k=o}^{n} c_k Q_k$. También pon-

ga $\langle f,g \rangle = \int_a^b f \, g \, w$ y sea $\{\overline{Q}_k\}$ el conjunto ortonormal de polinomios.

Ahora aplique la desigualdad de Bessel a $f - S_n$, notando que

$$f - S_n \perp \{Q_o, \ldots, Q_n\}$$

y

$$S_n \perp \{Q_{n+1}, \ldots\}$$

entonces

$$\sum_{k=n+1}^{\infty} \langle f, \overline{Q}_k \rangle^2 = \sum_{k=o}^{\infty} \langle f - S_n, \overline{Q}_k \rangle^2$$

$$\leq \langle f - S_n, f - S_n \rangle$$

$$\leq \langle f - P, f - P \rangle$$

$$\leq E_n^2 (f) \int_a^b w$$

Puesto que $\overline{Q}_k = Q_k / \langle Q_k, Q_k \rangle^{1/2}$ nosotros tenemos

$\langle f, \overline{Q}_k \rangle^2 = \langle f, Q_k \rangle^2 / \langle Q_k, Q_k \rangle = c_k^2 \langle Q_k, Q_k \rangle$. Esto prueba (ii)

Para la prueba (iii), observe que es trivial a menos que $\Sigma \gamma_k |c_k|$ sea -

convergente. En el caso último, las series $\Sigma\, c_k\, Q_k$ convergen uniformemente y representa f. Así

$$E_n(f) \leq ||f - S_n|| = || \sum_{k=n+1}^{\infty} c_k\, Q_k||$$

$$\leq \sum_{k=n+1}^{\infty} |c_k|\, ||Q_k|| = \sum_{k=n+1}^{\infty} |c_k|\, \gamma_k$$

La aplicación principal del teorema precedente es para expander las funciones en las series Tchebycheff, $f \sim \Sigma\, a_k\, T_k$, donde queremos decir mediante el símbolo $\sim$ unicamente que

$$a_k = \frac{2}{\pi} \int_0^{\pi} f(\cos\theta) \cos k\theta\, d\theta \text{ para } k \geq 1, \quad y \quad a_o = \frac{1}{\pi} \int_0^{\pi} f(\cos\theta)\, d\theta$$

no es del todo necesario que en éstas estimaciones la serie represente a f.

TEOREMA 5

Si $f \in C[-1,1]$ y si $f \sim \Sigma\, a_k\, T_k$, entonces

i) $E_n(f) \geq \frac{\pi}{4} \max\{ |a_{n+1}|, |a_{n+2}|, \dots \}$

ii) $E_n(f) \geq \sqrt{\frac{1}{2}(a_{n+1}^2 + a_{n+2}^2 + \dots)}$

iii) $E_n(f) \leq |a_{n+1}| + |a_{n+2}| + \dots$

PRUEBA

Debemos calcular las constantes α_k, β_k, γ_k que ocurren en el teorema anterior:

$$\int_{-1}^{1} T_k^2(x)(1 - x^2)^{-1/2}\, dx = \int_0^{\pi} \cos^2 k\theta\, d\theta = \frac{\pi}{2}$$

$$\int_{-1}^{1} |T_k(x)| \, (1 - x^2)^{-1/2} \, dx = \int_0^{\pi} |\cos k\theta| \, d\theta$$

$$= 2k \int_0^{\pi/2k} \cos k\theta \, d\theta = 2$$

$$\int_{-1}^{1} (1 - x^2)^{-1/2} \, dx = \int_0^{\pi} d\theta = \pi$$

Por lo tanto $\alpha_k = \frac{\pi}{4}$, $\beta_k = \frac{1}{2}$. Por supuesto que, $\gamma_k = 1$.

TEOREMA 6. [Rivlin, 1962 a]

Si $f \sim \Sigma\, a_k\, T_k$, entonces $\left| E_{n-1}(f) - |a_n| \right| \leq \sum_{k>n} |a_k|$

PRUEBA

La mitad de la desigualdad afirmada resulta del teorema anterior:

$$E_{n-1}(f) - |a_n| \leq \sum_{k>n} |a_k| - |a_n| = \sum_{k>n} |a_k|$$

Para la otra mitad nosotros debemos mostrar que

$$E_{n-1}(f) \geq |a_n| - \sum_{k>n} |a_k| \equiv \varepsilon$$

Esto es trivial si $\varepsilon \leq 0$, podemos por lo tanto asumir lo contrario. Entonces se da que

$$\left\| \sum_{k>n} a_k\, T_k \right\| \leq \sum_{k>n} |a_k| < |a_n| = \|a_n T_n\|$$

Consecuentemente la función $f - \sum_{k<n} a_k\, T_k = a_n\, T_n + \sum_{k>n} a_k\, T_k$ posee $n + 1$ puntos en los cuales toma alternativamente valores positivos y negativos, siendo estos puntos los extremos de T_n. Estos valores son al menos ε en magnitud, y por lo tanto mediante el teorema de La Vallée Poussin (T. 11 de Apéndice) $E_{n-1}(f) \geq \varepsilon$.

4. APROXIMACION DE CUADRADOS MINIMOS DISCRETA

Consideramos ahora un "seudo producto interno"

$$\langle f, g \rangle = \sum_{i=1}^{m} f(x_i)\, g(x_i)\, w(x_i)$$

en el cual los puntos x_i y los pesos $w(x_i)$ son prescritos y mantenidos fijos. Asumimos que f y g pertenecen a $C[a,b]$ y que $x_i \in [a,b]$. Llamamos a ésto un seudo producto interno porque allí existen funciones no cero f satisfaciendo $\langle f, f \rangle = 0$. Correspondiente a este seudo producto interno hay una seudo norma o seminorma.

$$||f|| = \langle f, f \rangle^{1/2}$$

y es razonable pregungar acerca de los problemas de aproximación relativos a ello. Resulta que algunos sistemas de funciones que fueron ortogonales con un producto interno integral son también ortogonales con respecto a un producto interno "discreto".

TEOREMA 1

Sea $\{\bar{Q}_0, \bar{Q}_1, \ldots\}$ el sistema de polinomios (los sub-índices denotan sus grados) el cual es ortonormal con respecto al producto interno:

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx.$$

Entonces el sistema $\{\bar{Q}_0, \bar{Q}_1, \ldots\}$ es también ortonormal con respecto al seudo producto interno

$$\langle f, g \rangle = \sum_{i=1}^{N} A_i\, f(x_i)\, g(x_i)$$

donde $N > n$, las x_i son las raices de $\bar{Q}_N$, y las A_i son los coeficientes

de cuadraturas Gaussianas.

PRUEBA

Para cada N hay una fórmula de integración Gaussiana (pág. 85),

$$\int_a^b f(x)\, w(x) \approx \sum_{i=1}^{N} A_i\, f(x_i)$$

la cual es exacta cuando f es un polinomio de grado < 2N. Aquí los x_i son las raíces de Q_N y los coeficientes son dados por las fórmulas de la pág. 89:

$$A_i = \frac{\phi_N(x_i)}{Q_N'(x_i)} = \frac{1}{Q_N'(x_i)} \int_a^b \frac{Q_N(x)}{x - x_i}\, w(x)\, dx$$

La positividad de los coeficientes A_i fue observada en el curso de probar el teorema de Stieltjes (sec. 1). Así el seudo producto interno - tiene la propiedad $\langle f, g \rangle \geq 0$. Tomando $k + m < N$ en la fórmula Gaussiana, tenemos

$$\delta_{km} = \int_a^b \overline{Q}_k(x)\, \overline{Q}_m(x)\, w(x)\, dx = \sum_{i=1}^{N} A_i\, \overline{Q}_k(x_i)\, \overline{Q}_m(x_i)$$

COROLARIO

Los polinomios de Tchebycheff tienen la siguiente propiedad, cuando $x_i = \cos\left[(2i - 1)\, \pi/2N\right]$ y $n + m < 2N$,

$$\frac{2}{N} \sum_{i=1}^{N} T_n(x_i)\, T_m(x_i) = \begin{cases} 0 & (n \neq m) \\ 1 & (n = m > 0) \\ 2 & (n = m = 0) \end{cases}$$

PRUEBA

La fórmula de cuadratura Gaussiana para este caso tiene los coefi-- cientes $A_i = \pi/N$. Como se observó en el problema 1, sec. 1, los polinomios Tchebycheff tienen la propiedad

$$\frac{2}{\pi} \int_{-1}^{1} T_n(x)\, T_m(x)(1-x^2)^{-1/2} dx = \begin{cases} 0 & (n \neq m) \\ 1 & (n = m > 0) \\ 2 & (n = m = 0) \end{cases}$$

de la cual resulta la ecuación deseada, sobre nuestra sustitución de la - suma Gaussiana para la integral.

TEOREMA 2

El polinomio P de grado $\leq n - 1$ que interpola a f en los ceros x1, ..., x_n de $\overline{Q}_n$ es dado por

$$P(x) = \sum_{k=0}^{n-1} a_k \overline{Q}_k \qquad a_k = \sum_{i=1}^{n} A_i \overline{Q}_k(x_i) f(x_i)$$

PRUEBA

Las ecuaciones mediante las cuales se determina P son

$$\sum_{k=0}^{n-1} a_k \overline{Q}_k(x_i) = f(x_i) \qquad (i = 1, \ldots, n)$$

Ahora multiplique ambos lados de esta ecuación por $A_i \overline{Q}_j(x_i)$ y sume para i = 1,..., n. Por el teorema anterior, obtenemos

$$a_j = \sum_{i=1}^{n} A_i \overline{Q}_j(x_i) f(x_{i.})$$

COROLARIO

El polinomio P de grado $\leq n-1$ el cual interpola a f en las raíces x_i de T_n es dado por las fórmulas

$$P = -\tfrac{1}{2}a_0 T_0 + \sum_{k=1}^{n-1} a_k T_k \qquad a_k = \frac{2}{n} \sum_{i=1}^{n} f(x_i)\, T_k(x_i)$$

En realidad, los polinomios de Tchebycheff tienen varias propiedades análogas del seno, coseno y funciones exponenciales. Una de las más importantes de estas propiedades está incluida en el siguiente teorema.

TEOREMA 3

Los polinomios Tchebycheff tienen la propiedad de ortogonalidad

$$\sum_{i=0}^{N}{}'' T_n(x_j)\, T_m(x_j) = \Delta(N, n, m)\, \frac{N}{2}$$

donde $x_i = \cos(i\pi/N)$, la doble prima significa que el primero y último término deben ser reducidos a la mitad y, $\Delta(N, n, m)$ denota el número de enteros en el conjunto $\{(n+m)/2N, (n-m)/2N\}$

PRUEBA

Puesto que $x_j = \cos(j\pi/N) = \cos[(2N-j)\pi/N] = x_{2N-j}$, tenemos

$$\sum_{j=0}^{N}{}'' T_n(x_j) T_m(x_j) = \frac{1}{2} \sum_{j=0}^{2N-1} T_n(x_j)\, T_m(x_j)$$

$$= \frac{1}{2} \sum_{j=0}^{2N-1} \cos\frac{nj\pi}{N} \cos\frac{mj\pi}{N}$$

Una aplicación de la identidad cos A cos B = $\frac{1}{2}\cos(A + B) + \frac{1}{2}\cos(A-B)$ produce

$$\frac{1}{4} \sum_{j=0}^{2N-1} \left(\cos \frac{n+m}{N} j\pi + \cos \frac{n-m}{N} j\pi \right)$$

la cual (a causa de la fórmula $e^{i\theta} = \cos\theta + i \operatorname{sen}\theta$) puede ser reconocida como la parte real de

$$\frac{1}{4} \sum_{j=0}^{2N-1} \left\{ \left[e^{(n+m)i\pi/N}\right]^j + \left[e^{(n-m)i\pi/N}\right]^j \right\}$$

Estas series geométricas son fácilmente sumadas usando la fórmula

$$\sum_{j=0}^{k-1} \lambda^j = \begin{cases} (1 - \lambda^k)(1 - \lambda)^{-1} & (\lambda \neq 1) \\ k & (\lambda = 1) \end{cases}$$

En efecto, si ni n-m, ni n+m es un múltiplo de 2N, la suma es

$$\frac{1 - e^{2(n+m)i\pi}}{1 - e^{(n+m)i\pi/N}} + \frac{1 - e^{2(n-m)i\pi}}{1 - e^{(n-m)i\pi/N}} = 0$$

Si uno de n-m ó n+m es un múltiplo de 2N, la suma es 2N, y si ambos son múltiplos de 2N, la suma es 4N.

TEOREMA 4

Si $f \in C[-1,1]$, entonces, con $x_i = \cos(i\pi/n)$,

$$E_{n-1}(f) \geq \frac{1}{n} \left| \sum_{i=0}^{n}{}'' (-1)^i f(x_i) \right|$$

(Las primas indican que el primero y último término en la suma deberán partirse)

PRUEBA

Denote P el polinomio de grano < n el cual representa la mejor aproximación de f en los n+1 puntos de $x_0, \ldots, x_n$.

Mediante el teorema de alternación (T.12 del Apéndice) P satisface las siguientes ecuaciones

$$(1) \qquad (-1)^i \lambda + P(x_i) = f(x_i)$$

Puesto que P es la mejor aproximación en $\{x_i\}$, tendremos

$$E_{n-1}(f) \geq \max_i |f(x_i) - P(x_i)| = |\lambda|$$

Nos queda entonces probar que

$$(2) \qquad n\lambda = \sum_{i=0}^{n}{}'' (-1)^i f(x_i)$$

Para hacer esto multiplique la ecuación (1) por $T_n(x_i)$, y luego aplique el operador $\sum''$ a ambos lados, así:

$$(3) \qquad \lambda \sum_{i=0}^{n}{}'' (-1)^i T_n(x_i) + \sum_{i=0}^{n}{}'' P(x_i) T_n(x_i) = \sum_{i=0}^{n}{}'' T_n(x_i) f(x_i)$$

Ahora P puede ser expresada como una combinación lineal de los polinomios de Tchebycheff $T_0, \ldots, T_{n-1}$. Por lo tanto mediante el teorema anterior, el término $\sum'' P(x_i) T_n(x_i)$ desaparece. Para los otros términos es necesario unicamente observar que $T_n(x_i) = \cos i\pi = (-1)^i$.

Es ahora una cosa fácil de establecer un teorema de [Bernstein, 1920] el cual provee una acotación inferior para $E_{n-1}(f)$ en aquellos casos cuando la serie de Tchebycheff para f converge absolutamente.

TEOREMA 5

Si $\Sigma|a_n| < \infty$ y $f = \Sigma\, a_n\, T_n$, entonces en $[-1, 1]$

$$E_{n-1}(f) \geq |a_n + a_{3n} + a_{5n} + \dots|$$

PRUEBA

Por el teorema anterior $E_{n-1}(f) \geq |\lambda|$ donde

$$\lambda = \frac{1}{n} \sum_{i=0}^{n''} (-1)^i f(x_i)$$

$$= \frac{1}{n} \sum_{i=0}^{n''} (-1)^i \sum_{k=0}^{\infty} a_k\, T_k(x_i)$$

$$= \frac{1}{n} \sum_{k=0}^{\infty} a_k \sum_{i=0}^{n''} T_n(x_i) T_k(x_i) \qquad (x_i = \cos\frac{i\pi}{n})$$

Hemos utilizado acá el hecho que $T_n(x_i) = (-1)^i$ y un teorema acerca del cambio del orden de la sumatoria. Ahora mediante el teorema 3, $\sum'' T_n(x_i)\, T_k(x_i) = n$ donde k es un múltiplo impar de n y de otra manera desaparece. Por lo tanto $\lambda = a_n + a_{3n} + \dots$

Hemos observado anteriormente que los polinomios de interpolación de Lagrange con nodos que han sido fijados con anticipación no proveen aproximaciones uniformes de precisión arbitraria a todas las funciones continuas. ¿Cuál es la situación para las normas que son diferentes a la norma uniforme? Un resultado positivo es el siguiente.

TEOREMA DE ERDÖS - TURAN

Sea Q_o, Q1,... el sistema de polinomios el cual es ortogonal en $[a,b]$ con función peso w. Para cada $f \in C[a,b]$ sea $L_n f$ la que denote el polinomio de grado $\leq n$ el cual interpola a f en los ceros de Q_{n+1}. Entonces $||L_n f - f||_w \to 0$. Es decir,

$$\int_a^b |(L_n f - f)(x)|^2 \, w(x) \, dx \to 0$$

PRUEBA

La fórmula de interpolación de Lagrange puede escribirse en la forma

$$(L_n f)(x) = \sum_{i=o}^{n} f(x_i) \, \ell_i(x) \qquad \ell_i(x) = \frac{Q_{n+1}(x)}{(x-x_i) \, Q'_{n+1}(x_i)}$$

De esto se da que $\ell_i \perp \ell_j$ para $i \neq j$; nosotros simplemente escribimos

$$\langle \ell_i, \ell_j \rangle = \frac{1}{Q'_{n+1}(x_i) \, Q'_{n+1}(x_j)} \int_a^b Q_{n+1}(x) \, \frac{Q_{n+1}(x)}{(x-x_i)(x-x_j)} \, w(x) \, dx$$

y observamos que la fracción bajo el signo integral es un polinomio de grado n - 1. Necesitaremos también la identidad

$$\Sigma \int \ell_i^2(x) \, w(x) \, dx = \int w(x) \, dx$$

Para probar esto, empezar por la ecuación $[\Sigma \ell_i(x)]^2 = 1$. Si nosotros multiplicamos por w(x) e integramos, el miembro izquierdo se simplifica (a través del uso de la propiedad de ortogonalidad $\ell_i \perp \ell_j$) a

$$\Sigma \int \ell_i^2 \, (x) \, w(x) \, dx.$$

Ahora para probar el teorema, dejemos que P_n denote el polinomio de

grado $\leq n$ el cual se aproxima mejor a f en la norma uniforme $||\cdot||_T$. Conforme a algún teorema de la página 94, $||P_n - f||_w \to 0$. Asi será suficiente para el propósito presente probar que $||L_n f - P_n||_w \to 0$. Ya que $L_n P_n = P_n$, tenemos, con la ayuda de los hechos anteriores,

$$\begin{aligned} ||L_n f - P_n||_w^2 &= ||L_n(f - P_n)||_w^2 \\ &= \int \{\Sigma[f(x_i) - P_n(x_i)]\ell_i(x)\}^2 w(x) \\ &= \Sigma\,[f(x_i) - P_n(x_i)]^2 \int \ell_i^2(x)\, w(x)\, dx \\ &\leq ||f - P_n||_T^2 \int w(x)\, dx \to 0 \end{aligned}$$

5. LOS TEOREMAS JACKSON

En la sección 3, nosotros derivamos algunos estimados de la cantidad

$$E_n(f) = \inf_{c_o,\ldots c_n} \sup_{-1\le x\le 1} |f(x) - \sum_{i=o}^{n} c_i\, x^i|$$

bajo el supuesto que f podría ser representada por una serie Tchebycheff, $\sum_{k=o}^{\infty} a_k\, T_k$. La única acotación superior en $E_n(f)$ obtenida fue dada por la desigualdad elemental

$$E_n(f) \le |a_{n+1}| + |a_{n+2}| + \ldots$$

Esto podría ser usado indirectamente en el caso de una función f - dos veces diferenciable para acotar $E_n(f)$ en términos de $||f''||$. Los teoremas de tal naturaleza, relacionando $E_n(f)$ a las propiedades de uniformidad de f, primero fueron dadas por Jackson en 1911. En los años intermedios los aspectos cuantitativos (pero no los cualitativos) de estos teoremas habían sido mejorados, por otros investigadores. En esta sección derivaremos un número de estos teoremas, señalando (pero no siempre probando) los resultados más posibles corrientemente conocidos.

Nuestro plan es obtener los estimados de $E_n(f)$ primero para la - aproximación por los polinomios trigonométricos. Dejemos que $C_{2\pi}$ represente el espacio de las funciones contínuas 2π - periódicas con la norma - suprema. Para $f \in C_{2\pi}$ oscribimos

$$E_n(f) = \inf_{a_k, b_k} \max_{\theta} |f(\theta) - \sum_{k=o}^{n} (a_k \cos k\theta + b_k \operatorname{sen} k\theta)|$$

El primer teorema de Jackson, en la forma mejorada debido a Favard

y Achieser - Krein, establece que $E_n(f) \leq (\frac{\pi}{2})(n+1)^{n-1}||f'||$ siendo la constante $(\frac{\pi}{2})(n+1)^{-1}$ la mejor posible. La prueba necesita 3 lemas.

LEMA 1

Si $k < n$, entonces $\int_0^{\pi} (\text{sen } kx)\ \text{sgn sen } nx\ dx = 0$

PRUEBA

Ya que el integrando es una función par, será suficiente probar

$$\int_{-\pi}^{\pi} (\text{sen } kx)\ \text{sgn sen } nx\ dx = 0$$

Puesto que sen kx es una combinación lineal de e^{ikx} y e^{-ikx}, será suficiente probar que cuando $|m| < n$,

$$\int_{-\pi}^{\pi} e^{imx}\ \text{sgn sen } nx\ dx = 0$$

Denote ahora la integral por I, y haga el cambio de la variable $x = y + \frac{\pi}{n}$. Así

$$I = \int_{-\pi-\pi/n}^{\pi-\pi/n} e^{im(y+\pi/n)}\ \text{sgn sen}(ny + \pi)\ dy$$

Ya que el integrando tiene período 2π, el intervalo de integración puede ser sustituido por $[-\pi, \pi]$. Por consiguiente

$$I = -e^{im\pi/n} \int_{-\pi}^{\pi} e^{imy}\ \text{sgn sen } ny\ dy = -e^{im\pi/n}\ I.$$

Ya que $|m| < n$, $m\pi/n$ no es un múltiplo impar de π, y $e^{im\pi/n} \neq -1$. Por consiguiente $I = 0$

LEMA 2

El valor mínimo de

$$\int_0^{\pi} \left| x - \sum_{k=1}^{n-1} \alpha_k \operatorname{sen} kx \right| \, dx$$

(para todas las posibles selecciones de α_k) es $\frac{\pi^2}{2n}$

PRUEBA

No importa la forma en que nosotros escojemos $\alpha_1, \ldots, \alpha_n$, nosotros podemos escribir utilizando la propiedad de ortogonalidad del lema 1,

$$\int_0^{\pi} \left| x - \sum_{k=1}^{n-1} \alpha_k \operatorname{sen} kx \right| dx \geq \left| \int_0^{\pi} \left(x - \sum_{k=1}^{n-1} \alpha_k \operatorname{sen} kx \right) \operatorname{sgn} \operatorname{sen} nx \, dx \right|$$

$$= \left| \int_0^{\pi} x \operatorname{sgn} \operatorname{sen} nx \, dx \right|$$

$$= \left| \sum_{k=0}^{n-1} (-1)^k \int_{k\pi/n}^{(k+1)\pi/n} x \, dx \right|$$

$$= \frac{\pi^2}{2n^2} \left| \sum_{k=0}^{n-1} (-1)^k (2k+1) \right| = \frac{\pi^2}{2n}$$

En el último paso nosotros hemos empleado una fórmula simple la cual puede probarse por inducción (ver prob. 2). Queda ahora ver si la acotación inferior $\pi^2/2n$ puede ser lograda para una escogitación particular de $\alpha_1, \ldots, \alpha_n$. Notando como las desigualdades se dan en el cálculo anterior, nosotros vemos que deberíamos hacer que la función $\phi(x) = x - \Sigma \alpha_k \operatorname{sen} kx$ cambie signo precisamente en los puntos de $(0, \pi)$ donde el sen nx cambia signo. Por lo tanto las ecuaciones, siguientes se satis

facerán para i = 1, ..., n-1.

$$\sum_{k=1}^{n-1} a_k \operatorname{sen} k x_i = x_i \qquad (x_i = \frac{i\pi}{n})$$

Que esto pueda hacerse es una consecuencia del hecho que

$$\{\operatorname{sen} x, \ldots, \operatorname{sen} (n-1) x\}$$

satisface la condición de Haar en $(0, \pi)$. Que la función resultante ϕ realmente cambie de signo en los puntos x_i será probado asumiendo lo contrario. Entonces ϕ' desaparece en cada intervalo (x_i, x_{i+1}), una vez en $(0, x1)$ y en una o más de los puntos x_i, para un total de n veces como mínimo. Pero ϕ' es de la forma $1 + \sum_{k=1}^{n-1} \beta_k \cos kx$ y puede desaparecer en no más de n-1 puntos de $(0, \pi)$.

Fijando n, definamos un operador L de la forma siguiente. Dado f una función contínua y 2π -periódica, colocar

$$(Lf)(x) = \frac{a_o}{2} + \sum_{k=1}^{n} A_k(a_k \cos kx + b_k \operatorname{sen} kx)$$

en el cual los coeficientes A_k quedan a nuestra disposición y el a_k y b_k son los coeficientes ordinarios de Fourier de f. La prueba del teorema de Jackson dependerá sobre una cierta selección de A_k la cual hace Lf una - buena aproximación a f. Nosotros requerimos una fórmula integral para Lf.

LEMA 3

Si f es 2π-periódica y si f' es contínua, entonces

$$(Lf - f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} \left[\frac{1}{2} t + \sum_{k=1}^{n} \frac{(-1)^k}{k} A_k \operatorname{sen} kt\right] f'(x + \pi - t)dt$$

PRUEBA

Si nosotros denotamos la expresión en corchetes por $\phi(t)$ y luego integramos el miembro derecho de la ecuación, por partes el resultado es

$$-\frac{1}{\pi}\ \phi(t)\ f(x+\pi-t)\ \Big|_{-\pi}^{\pi} + \frac{1}{\pi}\ \int_{-\pi}^{\pi}\ \phi'(t)f(x+\pi-t)dt$$

Utilizando las ecuaciones $\phi(\pm\pi) = \pm\frac{1}{2}\pi$ y $f(x) = f(x+2\pi)$, nosotros obtenemos

$$-f(x) + \frac{1}{\pi}\ \int_{-\pi}^{\pi}\left[\frac{1}{2} + \sum_{k=1}^{n}(-1)^k A_k \cos kt\right] f(x+\pi-t)\ dt.$$

Cambiando la variable $t \to x+\pi-s$, y usando $\cos k(x+\pi-s) = \cos k(x+\pi)\cos ks + \operatorname{sen} k(x+\pi)\operatorname{sen} ks = (-1)^k(\cos kx \cos ks + \operatorname{sen} kx \operatorname{sen} ks)$ nosotros llegamos a

$$-f(x) + \frac{1}{\pi}\ \int_{-\pi}^{\pi}\left[\frac{1}{2} + \sum_{k=1}^{n} A_k(\cos kx \cos ks + \operatorname{sen} kx \operatorname{sen} ks\right] f(s)\ ds.$$

Puesto que

$$a_k = \frac{1}{\pi}\ \int_{-\pi}^{\pi} f(s)\cos ks\ ds \quad y \quad b_k = \frac{1}{\pi}\ \int_{-\pi}^{\pi} f(s)\operatorname{sen} ks\ ds,$$

esta última expresión se vuelve inmediatamente $-f(x) + (Lf)(x)$.

TEOREMA I DE JACKSON.

Para toda función f 2π-periódica y con derivada contínua,

$$E_n(f) \leq \frac{\pi}{2(n+1)}\ ||f'||$$

y la constante $\pi/2(n+1)$ es la mejor posible.

PRUEBA

Del lema 3, no importa como seleccionemos $A_1, \ldots, A_n$,

$$E_n(f) \leq ||Lf - f||$$

$$\leq ||f'|| \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \frac{t}{2} + \sum_{k=1}^{n} \frac{(-1)^k}{k} A_k \operatorname{sen} kt \right| dt$$

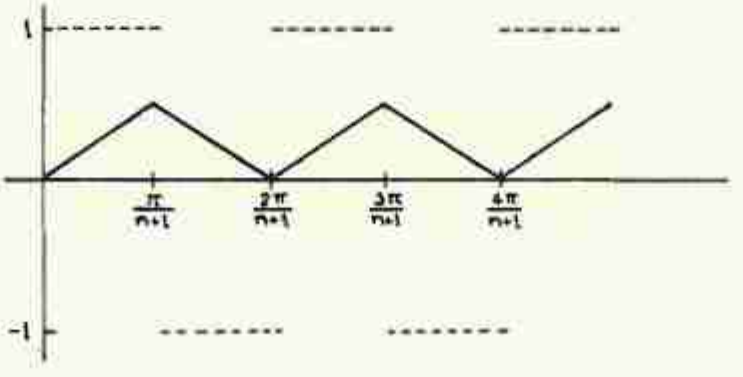
$$= ||f'|| \frac{1}{\pi} \int_{0}^{\pi} \left| t + \sum_{k=1}^{n} \frac{2(-1)^k}{k} A_k \operatorname{sen} kt \right| dt$$

Por el lema 2, existe una selección de los A_k para la cual esta última acotación superior se vuelve aquella del teorema.

Para probar que la constante es la mejor posible tenemos que excibir funciones para las cuales la acotación superior es casi alcanzada. A-fin de ver que propiedades deben tener tales funciones observe la fórmula en el lema 3. Los coeficientes A_k están seleccionados de tal forma que la expresión en corchete cambia de signo con el sen(n+1)t. Entonces el integral alcanza su magnitud máxima cuando f' cambia de signo con sen(n+1)t. Por lo tanto nosotros consideraremos funciones con derivadas contínuas - las cuales son cercanas a la función no diferenciable

$$f_o(x) = \int_0^x \operatorname{sgn} \operatorname{sen}(n+1)t \, dt$$

En el siguiente gráfico nosotros hemos trazado el integrando (con - línea cortada) y su integral f_o (con línea contínua)



La norma de f_o es claramente $\int_0^{\pi/(n+1)} 1\,dx = \pi/(n+1)$. El polinomio - trigonométrica de grado $\leq n$ el cual mejor se aproxima a f_o en $[0,2\pi)$ es la constante $\frac{1}{2}||f_o||$, puesto que el error tiene entonces 2n+2 puntos de alternación, es decir, los puntos $k\pi/(n+1)$ para $k=0,\ldots,2n+1$. Entonces $E_n(f_o) = \frac{1}{2}||f_o|| = \frac{\pi}{2}(n+1) = [\pi/2(n+1)]||f_o'||$

Puesto que esta función f_o es el límite de otras funciones las cuales tienen derivadas contínuas de norma 1, nosotros concluimos que la constante $\pi/2(n+1)$ es la mejor posible.

TEOREMA II DE JACKSON

Para todo $f \in C_{2\pi}$ el cual satisface $|f(x) - f(y)| \leq \lambda|x-y|$,

$$E_n(f) \leq \frac{\pi\lambda}{2(n+1)}$$

y la constante $\pi/2$ es la mejor posible.

PRUEBA

Fijando $\delta > 0$, definir $\phi(x) = \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} f(t)\,dt$. Entonces

$$|\phi'(x)| = \frac{1}{2\delta}\,|f(x+\delta) - f(x-\delta)| \leq \lambda$$

Consecuentemente, mediante el primer teorema de Jackson,

$$E_n(\phi) \leq \pi\lambda/2(n+1)$$

Además,

$$|\phi(x) - f(x)| \leq \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} |f(t) - f(x)|\,dt$$

$$\leq \frac{\lambda}{2\delta}\int_{x-\delta}^{x+\delta} |t - x|\,dt = \frac{\lambda}{2}\,\delta$$

Si P denota el polinomio trigonométrico de grado $\leq n$ el cual se aproxima mejor a ϕ, entonces

$$E_n(f) \leq ||f - P||$$
$$\leq ||f - \phi|| + ||\phi - P||$$
$$\leq \frac{\lambda}{2}\delta + \frac{\pi\lambda}{2(n+1)}$$

Ya que esto es verdadero para todo $\delta > 0$, es verdadero para $\delta = 0$, y esta es la desigualdad a ser probada. La constante $\pi/2$ es aquí la mejor posible porque es la mejor posible en la clase menor de las funciones con derivada contínua, y para tales funciones, $\lambda \leq ||f'||$. [En realidad para la función especial f_o considerada en el teorema anterior, se obtiene la cota $\pi\lambda/2(n+1)$.]

TEOREMA III DE JACKSON

Para toda $f \in C_{2\pi}$,

$$E_n(f) \leq w(\frac{\pi}{n+1})$$

donde w es el módulo de continuidad de f. El coeficiente 1 de $w[\pi/(n+1)]$ es el coeficiente mejor posible independiente de f y n.

En esta forma precisa del teorema de Jackson se debe a Korneicuk. Ya que la prueba es algo técnica, probaremos en cambio un resultado fácil pero débil, es decir,

$$E_n(f) \leq \frac{3}{2} w(\frac{\pi}{n+1})$$

PRUEBA

Empleando la función ϕ definida en la prueba del teorema anterior,

tenemos

$$|\phi'(x)| = \frac{1}{2\delta} \left| f(x+\delta) - f(x-\delta) \right| \leq \frac{1}{2\delta} w(2\delta)$$

Procediendo como lo hicimos anteriormente, obtenemos

$$|\phi(x) - f(x)| \leq w(\delta)$$

y entonces

$$E_n(f) \leq w(\delta) + \frac{\pi}{2(n+1)} \cdot \frac{1}{2\delta} w(2\delta)$$

$$\leq w(2\delta) \left[1 + \frac{\pi}{\delta 4(n+1)} \right]$$

Si se toma 2δ para ser $\pi/(n+1)$, ésto último se convierte en

$$\frac{2}{3} w \left[\pi/(n+1)\right]$$

EJEMPLO

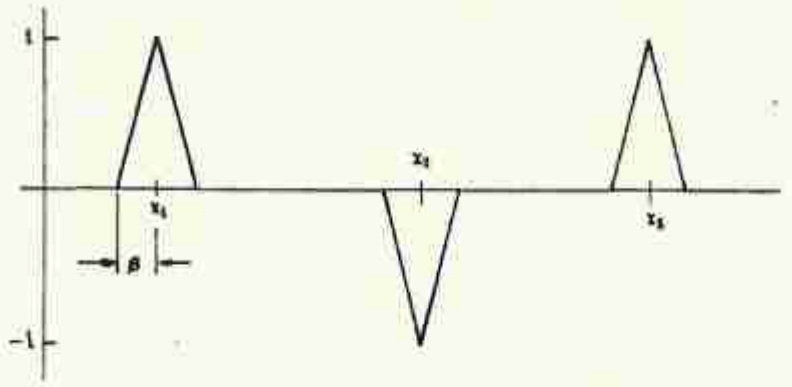
Mediante la excibición de una función especial, podemos mostrar que la desigualdad $E_n(f) \leq w\left[\pi/(n+1)\right]$es la mejor posible. Sea $0 < \varepsilon < 1/2$, y pongamos $h = \pi/(n+1)$. Seleccione $\beta \in (0; 2\varepsilon/(n+1)2)$, y ponga

$$x_i = ih - (n - i + 1)\beta \quad \text{para } i = 1, \ldots, n+1$$

Asi,

$$x_{i+1} - x_i = h + \beta > 0 \quad y \quad x_{n+1} = (n+1)h = \pi .$$

Defina ahora f por el gráfico siguiente. Coloque $f(-x) = f(x)$. Del gráfico observamos que $w(h) = 1$.



La prueba estará completa si podemos mostrar que

$$E_n(f) \geq (2n + 1)/(2n + 2) - \varepsilon$$

Sea $P(x) = [1/(n+1)]\ (\frac{1}{2} + \cos x + \ldots + \cos n\,x)$

Esta función (similar al núcleo de Dirichlet) se mostró en la pág. 102 que es la misma como

$$P(x) = \frac{\operatorname{sen}(n + \frac{1}{2})x}{2(n+1)\operatorname{sen}\frac{1}{2}x}$$

Observamos que $P(0) = (n + \frac{1}{2})/(n+1)$ y $P(ih) = (-1)^{i+1}/(2n + 2)$.

Además, $||P'|| \leq [1/(n+1)]\ (1 + 2 + \ldots + n) = n/2$. Por lo tanto

$$|P(x_i) - P(ih)| = |P'(\xi)|\ |x_i - ih| \leq (n/2)(n - i + 1)\beta \leq \varepsilon$$

Finalmente,

$$f(x_i) - P(x_i) = [f(x_i) - P(ih)] + [P(ih) - P(x_i)] = (-1)^{i+1} - (-1)^{i+1}/(2n + 2) + \delta_i, \quad \text{con } |\delta_i| \leq \varepsilon$$

Así $f - P$ toma en los $2n+2$ puntos $[-\pi, \pi)$ valores los cuales alternan en signo y son por lo menos $(2n + 1)/(2n + 2) - \varepsilon$ en magnitud, siendo el pun

to cero uno de éstos. Por el teorema de La Vallée Poussin,

$$E_n(f) \geq (2n + 1)/(2n + 2) - \varepsilon$$

TEOREMA IV DE JACKSON

Si $f \in C_{2\pi}$ y si f posee una k-ésima derivada contínua, entonces para $n > k$,

$$E_n(f) \leq \frac{\pi}{2}\left(\frac{1}{n+1}\right)^k ||f^{(k)}||$$

y el coeficiente $\pi/2$ es el coeficiente mejor posible independiente de f, k y n.

No daremos la prueba de este teorema, el cual depende de un análisis similar a aquel que conduce al teorema I, pero estaremos satisfechos con una prueba de la desigualdad más débil.

(1) $$E_n(f) \leq \left(\frac{\pi}{2n+2}\right)^k ||f^{(k)}||$$

PRUEBA

Denotemos por $e_n(f)$ el mínimo de $||f - P||$ cuando P se extiende sobre todos los polinomios trigonométricos de grado $\leq n$ con el período constante cero. La prueba de (1) consiste en establecer la sucesión de las desigualdades.

(2) $$\begin{aligned} E_n(f) &\leq \frac{\pi}{2n+2}\, e_n(f') \leq \left(\frac{\pi}{2n+2}\right)^2 e_n(f'') \\ &\leq \cdots \leq \left(\frac{\pi}{2n+2}\right)^{k-1} e_n(f^{(k-1)}) \\ &\leq \left(\frac{\pi}{2n+2}\right)^k ||f^{(k)}|| \end{aligned}$$

Para verificar la primera de éstas, sea p la mejor aproximación a f' libre de un término constante. Sea P un a integral indefinida de p. Entonces $||(f - P)'|| = ||f' - p|| = e_n(f')$. Por consiguiente por el teorema - I de Jackson,

$$E_n(f) = E_n(f - P) \leq \frac{\pi}{2n+2} ||(f - P)'|| = \frac{\pi}{2n+2} e_n(f')$$

Ahora en este argumento obtendríamos realmente

$$e_n(f) \leq \frac{\pi}{2n+2} e_n(f')$$

si las series de Fourier de f estuvieran libres de un término constante, ya que el operador L usado en la prueba del teorema I de Jackson produce un polinomio trigonométrico con la misma constante como en las series Fourier. En el caso de f' y todas las derivadas más altas la constante en las series Fourier es cero, debido a la periodicidad:

$$a_o = \frac{2}{\pi} \int_{-\pi}^{\pi} f'(x)\, dx = \frac{2}{\pi} [f(\pi) - f(-\pi)] = 0$$

Asi siempre tenemos $e_\nu(f^{(\nu)}) \leq [\pi/(2n + 2)]\ e_n(f^{(\nu+1)})$, para $\nu = 1, 2, \ldots$ La desigualdad final en (2) resulta del teorema I de Jackson, junto con la observación hecha acabada de hacer acerca del operador L.

Entre los corolarios a ser repetidos de los teoremas de Jackson está el teorema Dini-Lipschitz citada en la sec. 3. Nosotros lo establecemos aquí en términos de las series Fourier.

TEOREMA DINI-LIPSCHITZ

Si $f \in C_{2\pi}$ y si $w(\delta) \log \delta \to 0$ cuando $\delta \to 0$, entonces las series - Fourier de f converge uniformemente a f.

PRUEBA

La (n+1)-ésima suma parcial de la serie de Fourier de f fue mostrada en la pág. 101 que es de la forma

$$(S_n f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t + x) \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} \frac{1}{2} t} \, dt$$

De esto obtenemos inmediatamente

$$||S_n f|| \leq ||f|| \int_0^{\pi} \left| \frac{\operatorname{sen}(n + \frac{1}{2})t}{\pi \operatorname{sen} \frac{1}{2} t} \right| dt$$

La integración de la derecha produce un número conocido como la "n-ésima constante Lebesgue". Está acotada anteriormente por 3 + log n, - tal como veremos integrando separadamente $[0, \frac{1}{n}]$ y $[\frac{1}{n}, \pi]$ como sigue:

$$\frac{2}{\pi} \int_0^{1/n} \left| \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} \frac{1}{2} t} \right| dt = \frac{2}{\pi} \int_0^{1/n} \left| \frac{1}{2} + \cos t + \ldots + \cos nt \right| dt$$

$$\leq \frac{2}{\pi} \frac{1}{n} (\frac{1}{2} + n) < 1$$

Aquí hemos usado una identidad trigonométrica de la pág. 101
En el otro intervalo usamos el hecho que $\operatorname{sen}(\frac{t}{2}) \geq \frac{t}{\pi}$ [lo cual es evidente del gráfico del sen $(\frac{t}{2})$] para obtener

$$\frac{1}{\pi} \int_{1/n}^{\pi} \left| \frac{\operatorname{sen}(n + \frac{1}{2})t}{\operatorname{sen} \frac{1}{2} t} \right| dt \leq \frac{1}{\pi} \int_{1/n}^{\pi} \frac{1}{\frac{t}{\pi}} \, dt = \log \pi - \log \frac{1}{n}$$

$$< 2 + \log n.$$

Ahora sea P el polinomio trigonométrico de grado $\leq$ n el cual mejor se aproxima a f. Entonces mediante las observaciones anteriores y mediante el teorema III de Jackson.

$$||S_n f - f|| = ||S_n(f - P) - (f - P)||$$

$$\leq ||S_n(f - P)|| + ||f - P||$$

$$\leq (3 + \log n)||f - P|| + ||f - P||$$

$$= (4 + \log n)\, E_n(f)$$

$$< (4 + \log n)\ w(\tfrac{\pi}{n+1}) \to 0 \quad \text{cuando } n \to \infty$$

TEOREMA V DE JACKSON

Sea ahora $E_n(f)$ la que denote el error minimax al aproximar $f \in C[-1,1]$ mediante polinomios algebráicos de grado $\leq n$.

Entonces

i) $E_n(f) \leq w(\pi/(n+1))$

ii) $E_n(f) \leq [\pi\lambda/(2n+2)]$ si $|f(x) - f(y)| \leq \lambda |x - y|$

iii) $E_n(f) \leq (\pi/2)^k ||f^{(k)}||/[(n+1)(n)\ldots(n-k+2)]$ si $f^{(x)} \in C[-1, 1]$ y $n \geq k$.

PRUEBA

La función $g(\theta) = f(\cos\theta)$ es par contínua y 2π-periódica. Sus mejores aproximaciones mediante polinomios trigonométricos deberá ser por lo tanto par. Para ver esto sea P una mejor aproximación y sea $Q(\theta)=P(-\theta)$ Entonces

$$||Q - f|| = \max_{-\pi\leq\theta\leq\pi} |Q(\theta) - f(\theta)| = \max_{-\pi\leq\theta\leq\pi} |Q(-\theta) - f(-\theta)| = ||P - f||$$

de donde se da (por la unicidad de las mejores aproximaciones) que $P = Q$.

... y sea P = (P .. Entonces

Nosotros recordamos que todo polinomio trigonométrico par puede expresarse como un polinomio algebráico en la variable cos θ, e inversamente. Por lo tanto el error en la mejor aproximación de g mediante polinomos trigonométricos es el mismo que el error en la mejor aproximación de f por polinomios algebráicos:

$$\max_{-1\leq x\leq 1} |f(x) - P(x)| = \max_{-\pi\leq\theta\leq\pi} |f(\cos\theta) - P(\cos\theta)|$$

La aseveración (i) se dará ahora directamente mediante el teorema III de Jackson si nosotros podemos establecer que $w_g \leq w_f$. Que este es el caso puede verse utilizando el teorema del valor medio para obtener

$$|\cos\theta_1 - \cos\theta_2| = |-\operatorname{sen}\theta_3| \quad |\theta_1 - \theta_2| \leq |\theta_1 - \theta_2|$$

y escribiendo

$$w_g(\delta) = \max_{|\theta_1-\theta_2|\leq\delta} |g(\theta 1) - g(\theta 2)|$$

$$\leq \max_{|\cos\theta 1-\cos\theta 2|\leq\delta} |f(\cos\theta 1) - f(\cos\theta 2)| = w_f(\delta)$$

La afirmación (ii) se prueba en una forma similar según el teorema II de Jackson. Es unicamente necesario observar que si $|f(x)-f(y)| \leq \lambda|x-y|$ entonces $|g(\theta 1) - g(\theta 2)| \leq \lambda\,|\cos\theta 1 - \cos\theta 2| \leq \lambda|\theta_1 - \theta_2|$.

La prueba de la aseveración (iii) empieza con la desigualdad general

(1) $$E_n(f) \leq \frac{\pi}{2(n+1)} E_{n-1}(f')$$

A fin de verificar esto tome P_{n-1} que sea el polinomio de grado $\leq n - 1$ ei cual mejor se aproxima a f', y sea $P_n = \int P_{n-1}$. Entonces

$||(f - P_n)'|| = E_{n-1}(f')$. Consecuentemente, $f - P_n$ satisface una condición de Lipschitz con una constante $\lambda = E_{n-1}(f')$. Mediante la afirmación (ii) se da que $E_n(f - P_n) \leq \pi \lambda /(2n + 2)$, y esto es equivalente a la desigualdad (1).

Ahora aplique la desigualdad (1) k veces y luego use el hecho obvio $E_n(f) \leq ||f||$. El resultado es

$$E_n(f) \leq \frac{\pi}{2(n+1)} E_{n-1}(f') \leq \frac{\pi^2}{4(n+1)n} E_{n-2}(f'') \leq \cdots$$

$$\leq \left(\frac{\pi}{2}\right)^k \frac{1}{(n+1)(n)\ldots(n-k+2)} E_{n-k}(f^{(k)})$$

$$\leq \left(\frac{\pi}{2}\right)^k \frac{||f^{(k)}||}{(n+1)(n)\ldots(n-k+2)}$$

PROBLEMAS

1. Si $f \in C[-1,1]$ y $f \sim \sum a_k T_k$ entonces los polinomios $S_n f = \sum_{k=0}^{n} a_k T_k$ no son malos sustitutos para los polinomios de mejor aproximación a f. Específicamente, $E_n(f) \geq (4 + \log n)^{-1}||f-S_n f||$. Por lo tanto para todo n arriba de 400, nosotros podemos obtener a lo sumo una cifra decimal extra de seguridad al reemplazar $S_n f$ mediante el polinomio de mejor aproximación.

2. Probar que para $n = 0, 1, 2, \ldots,$ $\sum_{k=0}^{n} (-1)^k(2k + 1) = (-1)^n(n+1)$.

CAPITULO IV

APROXIMACION RACIONAL

1. LA EXISTENCIA DE LAS MEJORES APROXIMACIONES RACIONALES.

Consideramos el siguiente problema de aproximación. Nos son dadas una función $f \in C[a,b]$ y un par de enteros $n \geq 0$, $m \geq 0$. Nosotros buscamos el aproximar f mediante una función de la forma $R \equiv P/Q$, donde

$$P(x) = a_0 + a_1 x + \ldots + a_n x^n$$

$$Q(x) = b_0 + b_1 x + \ldots + b_m x^m$$

Nosotros podemos siempre tomar para nuestra función R una representación en la forma P/Q la cual es irreducible, es decir P y Q no tienen factores comunes más que constantes. Entonces, para que $R \equiv P/Q$ sea acotada en $[a,b]$ es necesario y suficiente que Q no tenga raíz en $[a,b]$. Por lo tanto, para aproximar funciones contínuas en la norma uniforme, no existe pérdida de generalidad al requerir que $Q(x) > 0$ en $[a,b]$. La familia resultante de funciones racionales se denota por $R_m^n[a,b]$:

$$R_m^n[a,b] = \{ \frac{P}{Q} \;/\; \partial P \leq n,\ \partial Q \leq m,\ Q(x) > 0 \text{ en } [a,b] \}$$

Aquí ∂P denota el grado de P, con el convenio que $\partial 0 = -\infty$. Nosotros adoptamos el convenio adicional que la representación irreducible de 0 es $\frac{0}{1}$.

Nosotros enfrentamos inmediatamente el problema de la existencia de la mejor aproximación en R_m^n. Una técnica general que sirvió para establecer los primeros problemas de existencia consistió en probar primero -

que el punto el cual nosotros encontramos perteneció a un conjunto compacto, el cual podría ser prescrito sobre razonamientos apriori. Por ejemplo, en el caso de los polinomios de grado $\leq n$, la mejor aproximación a f seguramente deberá descansar en el conjunto compacto

$$\{P/\ \partial P \leq n,\ ||f-P|| \leq ||f||\}$$

puesto que P debe proveer una aproximación a f, al menos tan buena como aquella proporcionada por cero.

La misma técnica no es efectiva en la presente circunstancia.

En efecto, el conjunto

$$\{R \in \mathbf{R}_m^n / ||R - f|| \leq ||f||\}$$

no es generalmente compacto. (La norma acá es la norma uniforme). Un ejemplo simple para sostener esta aseveración se da mediante la sucesión de funciones racionales.

$$R_k(x) = \frac{1}{kx+1} \qquad (k = 1, 2, 3, \ldots)$$

Sobre el intervalo $[\ 0,1\ ]$ éstas tienen la propiedad de $||R_k|| \leq 1$. Si fuese posible extraer de ellos una subsucesión convergente, entonces la función límite tendría que ser contínua. Pero esto no es posible ya que $R_k(0) = 1$ mientras que $R_k(x) \to 0$ cuando $x > 0$. A pesar de estas observaciones la compactación juega un rol crucial en el teorema de existencia.

TEOREMA DE EXISTENCIA

Para cada función $f \in C[\ a,b\ ]$ corresponde al menos una mejor aproximación racional proveniente de la clase $\mathbf{R}_m^n[\ a,b\ ]$.

PRUEBA

Sea $\delta = \text{dist}(f, R_m^n)$, y sea R_k una sucesión de elementos en R_m^n tal que $||R_k - f|| \to \delta$. Nosotros podemos escribir $R_k = P_k/Q_k$ donde $\partial P_k \leq n$, $\partial Q \leq m$, $||Q_k|| = 1$ y $Q(x) > 0$ en $[a,b]$. Pasando a una subsucesión si es necesario nosotros podemos asumir que $||R_k - f|| \leq \delta + 1$ para todo k. Consecuentemente $||R_k|| \leq ||R_k - f|| + ||f|| \leq \delta + 1 + ||f|| \equiv \theta$. Puesto que $|P_k(x)| = |Q_k(x)|\ |R_k(x)| \leq ||Q_k||\ ||R_k|| \leq \theta$, los pares (P_k, Q_k) están en el conjunto compacto definido mediante las desigualdades

$$||P|| \leq \theta \quad y \quad ||Q|| = 1$$

Pasando a una subsucesión si es necesario podemos asumir que $P_k \to P$ y $Q_k \to Q$. Claramente $||Q|| = 1$; por lo tanto pueden haber al máximo m puntos x_i donde $Q(x_i) = 0$. En todos los otros puntos $P(x)/Q(x)$ está bien definido y nosotros tenemos $P_k(x)/Q_k(x) \to P(x)/Q(x)$. Consecuentemente para estos puntos, $|P(x)/Q(x)| \leq \theta$, ó $|P(x)| \leq \theta|Q(x)|$. Por continuidad, esta última desigualdad es válida para todo x en $[a,b]$. Consecuentemente - cualquier cero de Q en $[a,b]$ es también un cero de P, y el factor lineal correspondiente a él puede ser cancelado de P y Q. Al quitar dicho factor lineal no hay molestia de la desigualdad previa y así nosotrospodemos repetir este proceso de cancelación hasta que Q se encuentre libre de ceros en $[a,b]$. Sea R el cual denote el elemento resultante de R_m^n Puesto que $R_k \to R$, $||R - f|| = \delta$.

Si uno intenta extender el teorema de existencia a las "funciones racionales generalizadas" de la forma

$$\frac{a_o g_o(x) + \ldots + a_n g_n(x)}{b_o h_o(x) + \ldots + b_m h_m(x)}$$

entonces algunas dificultades se encontraran debido a que la técnica de - factorización ya no está disponible. Uno puede sin embargo proceder de la siguiente forma.

DEFINICION

Sean todas las funciones g_i y h_i analíticas en $[a,b]$. Por lo tanto en cualquier punto $x \in [a,b]$ cada función posee una expansión de Taylor la cual representa a aquella función en un vecindario de x.

Denote R la familia de todas las funciones contínuas R en $[a,b]$ - las cuales satisfacen una ecuación de la forma

$$R(x) \Sigma b_i h_i(x) = \Sigma a_i g_i(x) \qquad (\Sigma |b_i| \neq 0)$$

TEOREMA

Cada función en $C[a,b]$ posee una mejor aproximación en R .

PRUEBA

Nosotros podemos asumir que el conjunto $\{h_o, \ldots, h_m\}$ es linealmente independiente, debido a que en el caso contrario $R = C[a,b]$, y el teorema es trivial. Seleccionemos elementos $R_k \in R$ de tal forma que $||f-R_k|| \to \delta = \text{dist}(f, R)$. Mediante la definición de R, existe para todo k, una - función P_k en el espacio lineal generado por $\{g_o, \ldots, g_n\}$ y una función Q_k en el espacio lineal generado por $\{h_o, \ldots, h_m\}$ tal que $R_k Q_k = P_k$ y $Q_k \neq 0$. No hay pérdida de generalidad al suponer que $||Q|| = 1$. Puesto -

que $||f - R_k|| \to \delta$, $||R_k||$ es acotada. Por lo tanto $||P_k||$ es acotada. - Por compactación, podemos asumir que $Q_k \to Q$ y $P_k \to P$. Claramente $||Q|| = 1$ Nosotros debemos ahora detenernos para probar que Q puede tener a lo sumo un número finito de ceros en $[a,b]$. El lector para quién este hecho es familiar deberá proceder al párrafo siguiente. Supóngase primero que Q desaparece idénticamente en un sub-intervalo $[\alpha,\beta]$. Nosotros podemos asumir que $\beta - \alpha$ es un máximo. Puesto que Q no es cero a través de $[a,b]$, nuestro sub-intervalo está propiamente contenido en $[a,b]$. Permitáconos decir por ejemplo que $\alpha > a$. Sea entonces la expansión de Taylor de Q en α, $Q(x) = \sum c_k(x - \alpha)^k$, y sea esta ecuación válida en un vecindario N de α. Puesto que $\beta - \alpha$ era máximal, existe un punto de N en el cual $Q(x) \neq 0$; - por lo tanto no desaparecen todos los coeficientes c_k. Sea c_ν el primer coeficiente distinto de cero. Entonces

$$Q(x) = (x - \alpha)^\nu \{ c_\nu + (x - \alpha)[c_{\nu+1} + c_{\nu+2}(x - \alpha) + \ldots] \}$$

De esta ecuación vemos que para todo x cercano a α, pero diferente de α, $Q(x) \neq 0$. En realidad, sea B una cota superior para el módulo de - la expresión en corchete, así como x varian en N. Si

$$0 < |x - \alpha|B < |c_\nu| \text{ y } x \in N,$$

entonces $Q(x) \neq 0$. Por lo tanto nosotros llegamos a la contradicción que- para algunos puntos en (α,β), $Q(x) \neq 0$. Ahora supóngase que Q posee un número infinito de ceros en $[a,b]$. Mediante la compactación podemos encontrar una sucesión convergente de ceros, es decir, $z_k \to z$. Puesto que Q no desaparece a través de ningún intervalo, la serie de Taylor de Q en z no es idénticamente cero. Procediendo como hicimos en el punto α, vemos -

que Q no puede desaparecer en todos los z_k. Ahora en cualquier punto x donde $Q(x) \neq 0$ podemos definir $R(x) = P(x)/Q(x)$, y R es contínua allí. Además $R(x) = \lim R_k(x)$ por lo tanto $|R(x) - f(x)| \leq \delta$: En cualquier punto z donde $Q(z) = 0$ nosotros escribimos las serie de Taylor

$$Q(x) = \sum_{k \geq \nu} c_k (x-z)^k \qquad y \qquad P(x) = \sum_{k \geq \mu} d_k (x-z)^k,$$

donde $c_\nu d_\mu \neq 0$. Puesto que $|R(x)|$ es acotada por $\delta + ||f||$ para todos los x cercanos a z pero diferentes de z, nosotros concluimos que $\mu \geq \nu$ Por lo tanto $P(x)/Q(x)$ está bien definida en un vecindario de z por la expresión

$$R(x) = \frac{d_\mu (x-z)^{\mu-\nu} + d_{\mu+1}(x-z)^{\mu-\nu+1} + \ldots}{c_\nu + c_{\nu+1}(x-z) + c_{\nu+2}(x-z)^2 + \ldots}$$

Claramente R es contínua en z y es un elemento de $\mathbf{R}$. Puesto que $|R(x) - f(x)| \leq \delta$ cuando $Q(x) \neq 0$, y R es contínua, esta desigualdad es verdadera también en los puntos donde $Q(x) = 0$. Por lo tanto $||R - f|| \leq \delta$

Nosotros concluímos esta sección con un teorema el cual garantiza la existencia de las mejores aproximaciones mediante funciones trigonométricas racionales

$$(1) \qquad R(\theta) = \frac{\sum_{j=0}^{n} (a_j \cos j\theta + b_j \operatorname{sen} j\theta)}{\sum_{j=0}^{m} (c_j \cos j\theta + d_j \operatorname{sen} j\theta)}$$

Si el intervalo se toma que sea $[-\pi, \pi]$ nosotros podemos siempre encontrar una mejor aproximación en la cual el denominador es estrictamente positivo. Esto es lo fundamental del asunto. Nosotros necesitamos-

un lema.

LEMA

Sean P y Q dos polinomios trigonométricos diferentes de cero con coeficiente reales tal que $|P(\theta)| \leq |Q(\theta)|$ para todo θ real. Si Q tiene un cero real, entonces existen allí polinomios trigonométricos distintos de ceros P* y Q* con coeficientes reales tal que $\partial P^* < \partial P$, $\partial Q^* < \partial Q^*$ y $P^*Q = PQ^*$.

PRUEBA

Es claro que el lema es verdadero para los polinomios algebráicos. La prueba para los polinomios trigonométricos será efectuada mapeando hacia y désde el dominio de los polinomios algebráicos. Definimos mediante la siguiente ecuación una transformación L_n del espacio de los polinomios trigonométricos reales de grado $\leq n$ en el espacio de los polinomios algebráicos reales de grado $\leq 2n$:

$$(2) \qquad (L_n f)(x) = (1 + x^2)^n f(2 \tan^{-1} x)$$

Que $L_n f$ es un polinomio algebráico de grado $\leq 2n$, siempre que f sea un polinomio trigonométrico de grado $\leq n$, puede ser observado comenzando con $\theta = 2 \tan^{-1} x$, de tal manera que

$\operatorname{sen} \theta = 2x(1 + x^2)^{-1}$ y $\cos x = (1 - x^2)(1 + x^2)^{-1}$.

Entonces $(1 + x2)^n \cos k\theta = (1 + x2)^n T_k(\cos \theta)$

$= (1+x^2)^n T_k [(1-x2)(1+x^2)^{-}1]$, el cual es un-

polinomio algebráico de grado $\leq 2n$. Aquí T_k denota el k-ésimo polinomio - Tchebycheff. Similarmente, usando los polinomios Tchebycheff U_k, tenemos

$$(1 + x^2)^n \operatorname{sen} k\theta = (1 + x^2)^n \operatorname{sen} \theta\, U_{k-1}(\cos \theta)$$

$$= (1 + x2)^n\, 2x(1 + x2)^{-1}\, U_{k-1}\left[(1-x2)(1+x^{2)^{-1}}\right]$$

El cual es un polinomio algebráico de grado $\leq 2n-1$. Es verdadero, pero- su verificación es dejada a los problemas, que L_n tiene un inverso dado por la fórmula

$$(L_n^{-1}f)(\theta) = (\cos \tfrac{\theta}{2})^{2n}\, f(\tan \tfrac{\theta}{2})$$

Ahora sean P y Q como en el lema. Podemos asumir sin pérdida de - generalidad que $Q(\pi) \neq 0$, ya que de otra manera hacemos un cambio de variable, $\theta \to \theta + \alpha$. Suponga que Q tiene una raíz $\theta_o \in (-\pi, \pi)$. Puesto que Q es periódica y contínua, tiene otra raíz en este mismo intervalo o tiene θ_o como una raíz doble. Es fácil de observar entonces, que L_mQ tiene- también dos raices reales. Ya que $|P(\theta)| \leq |Q(\theta)|$, obtenemos

$$|(L_nP)(x)| \leq (1 + x^2)^{n-m}\, |(L_mQ)(x)|$$

Esto muestra que cualesquiera de las raices reales de L_mQ están presen-- tes, con multiplicidades al menos tan grande como, en L_nP. Así L_mQ y L_nP comparten un factor cuadrático correspondientes a dos raíces reales. Los polinomios que resultan de remover este factor cuadrático pueden ser denotados por $L_{n-1}P^*$ y $L_{m-1}Q^*$, donde P^* y Q^* son polinomios trigonométricos de grados $< n$ y $< m$, respectivamente.

Este argumento requiere la invertibilidad de los operadores L_{n-1} y $L_{m-}1$. Que $PQ^* = P^*Q$ resulta del hecho que $L_nP/L_mQ = L_{n-}1P^*/L_{m-}1Q^*$ y de la ecuación (2)

TEOREMA

A cada $f \in C[-\pi, \pi]$ corresponde una función trigonométrica racional de la forma anterior (1) la cual se aproxima mejor a f. Con ninguna pérdida de generalidad puede asumirse que esta función tiene un denominador positivo.

PRUEBA

Por el teorema que precede al lema, existe una función trigonométrica racional, P/Q, de mejor aproximación, pero en puntos donde el denominador desaparece debe ser definido como un límite. Si esto ocurre, una desigualdad $|P(\theta)| \leq |k\ Q(\theta)|$ es claramente válida. De esta forma podemos aplicar el lema (tal vez repetidamente) para obtener otros polinomios trigonométricos P* y Q* tal que $\partial P^* < \partial P$, $\partial Q^* < \partial Q$, $P/Q = P^*/Q^*$, y $Q^*(\theta) > 0$ en $[-\pi, \pi]$

PROBLEMA

1. Verificar la fórmula para L_n^{-1} tal como fue dada en la prueba del lema. Probar también que $\partial L_n P = 2n - k$, donde k es la multiplicidad de π como una raíz de P. Probar que L_n es lineal.

2. LA CARACTERIZACION DE LAS MEJORES APROXIMACIONES

Nuestro objetivo inmediato es establecer para las aproximaciones racionales generalizadas el análogo del teorema de caracterización, dado en el teorema 13 del Apéndice. Esto a su vez conducirá a una caracterización de las mejores aproximaciones por medio de las alternaciones en la curva de error. En esta sección adoptamos un ajuste el cual es mucho más general que el ajuste que fue requerido por el teorema de existencia de la sección anterior. Supongamos ahora que dos subespacios de dimensión finita **R** y **Q** han sido prescritos en $C[X]$. El conjunto X puede ser cualquier espacio métrico compacto, aun cuando más tarde requeriremos que X sea un intervalo Se asume que **Q** contiene por lo menos una función la cual es positiva a través de X. Nuestra familia de aproximación es entonces la clase **R** de todas las funciones $R = P/Q$, donde $P \in \mathbf{P}$, $Q \in \mathbf{Q}$, y $Q(x) > 0$ en X. Tales funciones R serán llamadas funciones racionales generalizadas.

Si f es un elemento dado de $C[X]$, pueda que exista o puede no existir en **R** un elemento de mejor aproximación a f. Un caso en el cual la existencia ya ha sido probada es aquel en el cual **P** consiste de todos los polinomios de grado $\leq n$, **Q** consiste de todos los polinomios de grado $\leq m$, y X es un intervalo. Otro caso es aquel en el cual **Q** tiene dimensión 1 y **P** es arbitraria, siendo éste el problema puramente lineal. Aun en el caso general, donde un teorema de existencia está faltando, seremos capaces de caracterizar las mejores aproximaciones desde **R**. Dado un elemento fijo R en **R**, escribiremos

$$\mathbf{P} + R\mathbf{Q} = \{P + RQ/P \in \mathbf{P} \text{ y } Q \in \mathbf{Q}\}$$

Este es un subespacio lineal de $C[a,b]$. Si $(g_1,\ldots, g_n)$ es una base para P y si $\{h1,\ldots,h_m\}$ es una base para Q, entonces P + R Q es generado por

$$\{g_1,\ldots, g_n, Rh_1,\ldots, Rh_m\}$$

Aun cuando R≠0, el último no es una base. En efecto, si

$$R = \Sigma a_i g_i / \Sigma b_i h_i,$$

entonces tenemos la dependencia lineal

$$\Sigma a_i g_i - \Sigma b_i Rh_i = 0$$

Asi P + R Q puede tener dimensión al máximo de n + m - 1.

TEOREMA DE CARACTERIZACION

Un elemento $R \in R$ es una mejor aproximación a $f \notin R$ si y sólo si ningún elemento $\phi \in P + R Q$ tiene los mismos signos como f - R en el conjunto de los puntos críticos

$$Y = \{ y/|f(y) - R(y)| = ||f - R||\}$$

PRUEBA

Si R no es una mejor aproximación a f, seleccione una mejor, $R^* = P^*/Q^* \in R$. Ponga $\phi = Q^*(R^* - R)$. Este es un elemento de P + R Q. Además, si $y \in Y$ y si $\sigma(y) = Sgn(f - R)(y)$, entonces $\sigma(y)(f - R^*)(y) \leq ||f - R^*|| < ||f - R|| = \sigma(y)(f - R)(y)$ de donde $\sigma(y)(R^* - R)(y) > 0$ y $\sigma(y)\phi(y) > 0$

Para el inverso, dejemos que ϕ concuerde en signo con f - R en Y. Escriba $\phi = P_o + RQ_o$, $R = P/Q$, y

$$R_\lambda = \frac{P + \lambda P_o}{Q - \lambda Q_o}$$

El resto de la prueba está dedicado a mostrar como seleccionar λ de-tal forma que $||f - R_\lambda|| < ||f - R||$. Defina

$$\delta = \inf_{x \in Y} \sigma(x)\phi(x)$$

Por la continuidad y compactación, $\delta > 0$. Dejemos que $e = f - R$, y defina los conjuntos

$$X_1 = \{x \in X / \sigma(x)\phi(x) > \tfrac{1}{2}\delta \text{ y } |e(x)| > \tfrac{1}{2}||e||\}$$

$$X_2 = X \sim X_1$$

Es claro que X_1 contiene a Y y que X_2 es un conjunto compacto que no tiene puntos de Y. Por lo tanto existe un número μ que satisface las desigualdades

$$|e(x)| \leq \mu < ||e|| \qquad (x \in X_2)$$

A fin de determinar las restricciones adecuadas sobre λ , debemos realizar algunos cálculos. Primero, para $x \in X_2$, tenemos

$$\begin{aligned} |f(x) - R_\lambda(x)| &\leq |f(x) - R(x)| + |R(x) - R_\lambda(x)| \\ &\leq \mu + ||R - R_\lambda|| \end{aligned}$$

Puesto que $||R - R_\lambda|| \to 0$ cuando $\lambda \to 0$, este último término es menor que $||e||$ para todos los λ suficientemente pequeños. Ahora nosotros tomamos un λ tan pequeño que $f(x) - R_\lambda(x)$ tenga el mismo signo que $f(x) - R(x)$ en X_1. Por lo tanto para $x \in X_1$

$$\begin{aligned} |f(x) - R_\lambda(x)| &= \sigma(x)(f - R)(x) + \sigma(x)(R - R_\lambda)(x) \\ &\leq ||e|| - \frac{\lambda\sigma(x)\phi(x)}{(Q - \lambda Q_o)(x)} \end{aligned}$$

$$\leq ||e|| - \frac{\lambda\,\delta}{2||Q-\lambda Q_0||} < ||e||$$

En este argumento es necesario restringir λ a valores positivos pequeños tales que $Q - \lambda Q_0$ es positivo a través de X.

El teorema de caracterización recien probado puede ser enunciado en otras formas equivalentes. Por ejemplo:

TEOREMA

Un elemento $R \in \mathbf{R}$ es una mejor aproximación a f si y sólo si existen puntos $x_i \in X$ y escalares $\lambda_i \neq 0$ tal que

i) $f(x_i) - R(x_i) = (\text{Sgn } \lambda_i)\ ||f - R||$

ii) $\Sigma \lambda_i \phi(x_i) = 0$ para cada $\phi \in \mathbf{P} + R\,\mathbf{Q}$

TEOREMA

Un elemento $R \in R$ es una mejor aproximación a f si y sólo si el origen del espacio n está situada en la cápsula convexa del conjunto

$$\{\sigma(x)\hat{x} / |f(x) - R(x)| = ||f - R||\}$$

donde $\sigma(x) = \text{Sgn}[\, f(x) - R(x)\,]$, $\hat{x} = [\phi_1(x),\ldots,\phi_n(x)]$, y $\{\phi_1,\ldots,\phi_n\}$ es cualquier base para $\mathbf{P} + R\,\mathbf{Q}$.

Hablaremos de un subespacio de Haar de $C[a,b]$ como un subespacio dimensional-finito el cual tiene una base que satisface la condición de Haar. Por lo tanto M es un subespacio de Haar si existe una base $\{g1,\ldots, g_n\}$ para M tal que cada determinante

$$\begin{vmatrix} g_1(x_1) & \dots & g_n(x_1) \\ \dots & \dots & \dots \\ g_1(x_n) & \dots & g_n(x_n) \end{vmatrix}$$

(formada con puntos diferentes $a \leq x_i \leq b$) es diferente a cero.

Que esta propiedad de M es independiente de la base se sigue inmediatamente: M es un subespacio de Haar de dimensión n si y sólo si cero es la única función en M la cual tiene n o más raíces en [a,b]

Si el subespacio M es o no un subespacio de Haar, el número de cambios en el signo el cual pueden poseer sus miembros tiene al menos una acotación superior la cual depende unicamente de M.

Sea esto denotado por $\nu(M) - 1$. Nosotros admitimos la posibilidad que para algunos subespacios M, $\nu(M)$ puede ser $+\infty$. Luego denotemos por $\delta(M)$ la dimensión de M. Se sigue que todo subespacio de Haar M satisface la ecuación $\delta(M) = \nu(M)$. Finalmente, en cualquier subespacio M nosotros podemos buscar subespacios de Haar. Sea $\eta(M)$ la dimensión máxima entre estos. Entonces M por si mismo es un subespacio de Haar si y sólo si $\delta(M) = \eta(M)$. En resumen,

$\delta(M)$ = dimensión de M

$\nu(M)$ = 1 + el número máximo de variaciones en signos poseidos por miembros de M.

$\eta(M)$ = max {$\delta(H)$/H es un subespacio Haar de M}

Dado un elemento R de **R** formamos de nuevo el subespacio **P** + R **Q** . Los indices $\nu(\mathbf{P} + R\,\mathbf{Q})$ y $\eta(\mathbf{P} + R\,\mathbf{Q})$ dependen ahora unicamente de R. Recalcamos que se dice que una función e tiene k alternaciones si existen

puntos $x1 < \ldots < x_k$ talque $e(x_i) = (-1)^i\lambda$, con $|\lambda| = ||e||$.

TEOREMA DE ALTERNACION

Si la función error $e = f - R$ tiene al menos $1 + \nu(P + R\,Q)$ alternaciones, entonces R es una mejor aproximación a f desde $\mathbf{R}$. Si R es una mejor aproximación a f, entonces e tiene al menos $1 + \eta(P + R\,Q)$ alternaciones.

PRUEBA

Si R no es una mejor aproximación a f, entonces mediante el teorema de caracterización, podemos encontrar $\phi \in P + R\,Q$ tal que

$$|e(x)| = ||e|| \Rightarrow \phi(x)\, e(x) > 0$$

Esto muestra que si e alterna cada k veces, entonces tiene k - 1 variaciones en signo. Por lo tanto e no puede alternar mas que $\nu(P + R\,Q)$ veces.

Suponga ahora que R es una mejor aproximación a f. En $P + R\,Q$ podemos seleccionar un subespacio de Haar M de dimensión $n = \eta(P + R\,Q)$. Dejemos que $Y = \{x/|e(x)| = ||e||\}$

Por el teorema de caracterización, no hay ningún $\phi \in M$ tal que $\phi(x)\, e(x) > 0$ en Y. Si $\{\phi 1, \ldots, \phi_n\}$ es una base para M, entonces el sistema de desigualdades

$$e(x)\ \Sigma\, c_i\phi_i(x) > 0 \qquad (x \in Y)$$

es inconsistente, y así el origen del espacio n-dimensional está situado -

en la cápsula convexa del conjunto establecido

$$\{e(x)\hat{x}/x \in Y\}$$

donde $\hat{x}$ denota $[\phi_1(x),\ldots,\phi_n(x)]$. Mediante el teorema de Carathéodory, y por la condición de Haar el origen está situado en la cápsula convexa - de algún conjunto de precisamente d e n + 1 de tales puntos, $e(x_i)\hat{x}_i$. Por- el lema 1 del Apéndice, los números $e(x_i)$ deben alternar en signo si $x_1 < x_2 < \ldots$. Por lo tanto e alternan+ 1 veces.

El teorema anterior da una caracterización completa de las mejores - aproximaciones racionales generalizadas solamente cuando los dos índices - n y v son los mismos para el subespacio $\mathbf{P} + R\,\mathbf{Q}$. Afortunadamente esto se - vuelve verdadero para las aproximaciones racionales ordinarias, de conformidad al siguiente lema.

LEMA

Sean **P** y **Q** los espacios de los polinomios de grado $\leq n$ y $\leq m$, respectivamente. Dejemos que $R = P/Q$, con $P \in \mathbf{P}$, $Q \in \mathbf{Q}$, $Q > 0$ en $[a,b]$,y P/Q irreducible. Entonces $\mathbf{P} + R\,\mathbf{Q}$ es un subespacio de Haar en $C[a,b]$ de dimensión $1 + \max\{n + \partial Q, m + \partial P\}$.

PRUEBA

Comenzamos por mostrar que la dimensión de $\mathbf{P} + R\,\mathbf{Q}$ es

$$k \equiv 1 + \max\{n + \partial Q, m + \partial P\}.$$

Si $R = 0$, entonces por nuestras convenciones, $P = 0$, $Q = 1$, y $\partial P = -\infty$ de tal forma que $k = 1 + n$, y esta es la dimensión de **P**. En el otro caso - $(R \neq 0)$, usamos la ecuación

$$\delta(P + R\,Q) = \delta(P) + \delta(R\,Q) - \delta(P \cap R\,Q)$$

La dimensión de P es n + 1, y la dimensión de R Q es n + 1.

Ahora $R\,Q = \{(P/Q)Q_1 / \partial Q1 \leq m\}$, y un elemento (P/Q)Q1 pertenecerá también a P si y sólo si Q divide a Q1, dejando un cociente de grado $\leq n - \partial P$. En este caso, Q_1 debe ser de la forma QQ_2 con $\partial Q_2 \leq n - \partial P$. Puesto que $\partial Q_1 \leq m$, debemos también tener $\partial Q_2 \leq m - \partial Q$. Asi $\delta(P \cap R\,Q) = 1 + \min\{ m - \partial Q, n-\partial P\}$ $= m + n + 2 - k$, de donde $\delta(P + R\,Q) = k$.

Ahora para probar que P + R Q es un subespacio de Haar solamente necesitamos establecer que sus elementos no triviales puedan tener a lo sumo k-1 raíces en $[\,a,b\,]$. Si uno de sus elementos, digamos $P_1 + R\,Q1$, tiene k raíces, entonces $P_1Q + P\,Q_1$ también tiene k raíces. Pero esto no es posible ya que este polinomio es de grado a lo sumo

$$\max\{n + \partial Q, m + \partial P\} = k - 1$$

COROLARIO

Para que la función racional irreducible P/Q sea una mejor aproximación a f de la clase $R_m^n[a,b\,]$, es necesario y suficiente que el error - tenga por lo menos $2 + \max\{ n + \partial Q, m + \partial P \}$ alternaciones.

3. UNICIDAD; CONTINUIDAD DE LOS OPERADORES DE MEJOR APROXIMACION

Retenemos la formación de conjuntos de la sección previa. Asi nuestra familia de aproximación es el conjunto R de todas las funciones $R = P/Q$, - donde P varía en un subespacio dimensional-finito P de $C[a,b]$, y Q varía en otro subespacio dimensional - finito Q, pero sujeto a la restricción - $Q(x) > 0$ en $[a,b]$. Dado $R \in R$, formamos el subespacio $P + RQ$ consistiendo de todos los $P1 + RQ_1$ con $P1 \in P$ y $Q_1 \in Q$. Los índices η, δ y ν de la pág. 155 jugaran de nuevo un papel. El primer resultado es un lema el cual fortalece el teorema de caracterización en el caso que $P + RQ$ sea un subespacio de Haar.

LEMA 1

Sea R una mejor aproximación en P a una función $f \notin R$. Si $P + RQ$ es un subespacio de Haar entonces cero es el único elemento ϕ de $P + RQ$ teniendo la propiedad $\phi(y)(f-R)(y) \geq 0$ para todo y en el conjunto de puntos - críticos, $Y = \{y/|f(y) - R(y)| = ||f - R||\}$

PRUEBA

Sean $\{\phi 1, \ldots, \phi_n\}$ una base para $P + RQ$. Asi como en el teorema de-alternación (sec. 2) inferimos que el origen del espacio n está situado en la cápsula convexa del conjunto

$$\{e(x)\hat{x}/x \in Y\}$$

donde $\hat{x} = [\phi_1(x), \ldots, \phi_n(x)]$ y $e = f - R$. Escribamos

$$0 = \sum_{i=0}^{k} \theta_i e(x_i)\hat{x}_i,$$

con $x_i \in Y$, y $\theta_i > 0$. Mediante la condición de Haar, $k \geq n$. Si ϕ es cualquier elemento distinto de cero de $P + R\,Q$, entonces

$$0 = \sum_{i=0}^{k} \theta_i e(x_i)\phi(x_i).$$

Mediante las condiciones de Haar, a lo sumo n - 1 de los números $e(x_i)\phi(x_i)$ pueden desaparecer. Por lo tanto al menos uno de estos es positivo y al menos uno es negativo.

TEOREMA DE LA UNICIDAD

Sea R una mejor aproximación de **R** a la función f. Si $P + R\,Q$ es un subespacio Haar, entonces R es único.

PRUEBA

Suponga al contrario que $R_o \equiv P_o/Q_o$ es otra mejor aproximación. La función $\phi = Q_o(R_o - R)$ pertenece a $P + R\,Q$, y de identidad

$$\phi = Q_o(R_o - R) = Q_o\left[(f - R) - (f - R_o)\right]$$

se da para todos los puntos críticos y (de la función f - R) debemos tener $\phi(y)(f - R)(y) \geq 0$. Puesto que $f \notin R$, el lema 1 puede ser aplicado, con la conclusión que $\phi = 0$ y $R = R_o$.

COROLARIO (Unicidad para la aproximación racional ordinaria)

Las mejores aproximaciones en $R_m^n[a,b]$ son siempre únicas.

PRUEBA

En la sección precedente, hemos establecido que en el caso de aproxi

mación racional ordinaria, $P + R\,Q$ es siempre un subespacio de Haar. La - unicidad se da entonces como resultado del teorema.

Deberá notarse que aún si una mejor aproximación en R es única, su representación P/Q nunca es única. Podemos por ejemplo multiplicar el numerador y el denominador por cualquier escalar positivo. En algunos casos existe la posibilidad de multiplicar numerador y denominador por una función - diferente a una constante. Por ejemplo, si $P/Q \in R_m^n[a,b]$, si $\partial P < n$, y $\partial Q < m$, entonces P y Q pueden multiplicarse por cualquier factor $x - c$ con $c < a$. Antes de dar el teorema fuerte de unicidad en el caso de aproximación racional, es conveniente extraer un lema el cual incorpora ciertos - argumentos de dimensión.

LEMA 2

Sea $R^* = P^*/Q^*$ un elemento de R tal que $\delta(P + R^* Q) = \delta(P) + \delta(Q) - 1$. Si $P \in P$, $Q \in Q$, $||P|| + ||Q|| = ||P^*|| + ||Q^*||$, $P = R^*Q$, y $Q(x) > 0$ en $|\,a,b\,|$, entonces $P = P^*$ y $Q = Q^*$.

PRUEBA

Si $R^* = 0$, entonces $P^* = 0$ y $P = 0$. Además $\delta(Q) = 1$. Por lo tanto - $||Q|| = ||Q^*||$ y $Q(x)\,Q^*(x) \geq 0$, se da que $Q = Q^*$.

Si $R^* \neq 0$, entonces las ecuaciones $P = R^* Q$ y $P^* = R^* Q^*$ muestre que P y P^* son elementos diferentes de cero de $P \cap R^* Q$. Sin embargo la desigualdad

$$\delta(P + R^* Q) \leq \delta(P) + \delta(Q) - \delta(P \cap R^* Q)$$

muestra que $\delta(P \cap R^* Q) \leq 1$. Por lo tanto P es un escalar múltiplo de P* Mediante las condiciones restantes vemos que $P = P^*$ y $Q = Q^*$.

TEOREMA FUERTE DE UNICIDAD

Sea R* una mejor aproximación en R a f. Si $\eta(P + R^* Q) = \delta(P) + \delta(Q) - 1$, entonces existe una constante $\gamma > 0$ tal que para todo $R \in R$,

$$||f - R|| \geq ||f - R^*|| + \gamma ||R - R^*||$$

PRUEBA

El teorema es trivial en el caso que $f \in R$. Nosotros asumimos lo contrario. Para $R \in R$ y $R \neq R^*$ definamos

$$\gamma(R) = \frac{||f - R|| - ||f - R^*||}{||R - R^*||}$$

Nuestra tarea es probar que $\gamma(R)$ es acotada lejos del cero. Supónga se al contrario que es posible encontrarse una sucesión $R_k \in R$ tal que $R_k \neq R^*$ y $\gamma(R_k) \to 0$. Ponga $R_k = P_k/Q_k$ con $P_k \in P$ y $Q_k \in Q$. Podemos asumir que $||P_k|| + ||Q_k|| = 1$. De igual manera, si $R^* = P^*/Q^*$, nosotros podemos asumir que $||P^*|| + ||Q^*|| = 1$. Por compactación podemos asumir que $P_k \to P$ y $Q_k \to Q$. Puesto que $\gamma(R_k) \to 0$, $||R_k||$ y $||R_k - R^*||$ permaneoen acotados.

Nosotros mostramos primero que $P = R^* Q$. Dejemos que

$$\sigma(x) = \text{Sgn}(f - R^*)(x),$$

y denote y un punto crítico arbitrario de $f - R^*$; es decir,

$$y \in Y = \{x / |(f - R^*)(x)| = ||f - R^*||\}$$

Luego se da que

$$(1)\qquad \gamma(R_k)\,\|R^* - R_k\| = \|f - R_k\| - \|f - R^*\|$$

$$\geq \sigma(y)(f - R_k)(y) - \sigma(y)(f - R^*)(y)$$

$$= \sigma(y)(R^* - R_k)(y)$$

$$= \frac{\sigma(y)(R^*Q_k - P_k)(y)}{Q_k(y)}$$

Pasando al limite nosotros obtenemos

$$\sigma(y)(P - R^*Q)(y) \geq 0 \qquad (y \in Y)$$

Mediante el lema 1, esto implica que $P = R^*Q$. (Observe que $P + R^*Q$ es un subespacio de Haar puesto que los dos índices n y δ son lo mismo para ello). Mediante el lema 2, se da que $P = P^*$ y $Q = Q^*$.

Puesto que $Q^*(x) > 0$ en $[a,b]$, podemos pasar a una subsucesión tal que para algún $\varepsilon > 0$ y para todo k y x, $Q_k(x) \geq \varepsilon$. Ahora definamos

$$c = \inf_{\substack{\phi \in P+R^*Q \\ \|\phi\|=1}} \max_{y \in Y} \sigma(y)\phi(y)$$

El lema 1 implica que $c > 0$. De la definición de c y de la deslgualdad (1) se da que existe un $y \in Y$ con la propiedad

$$\gamma(R_k)\,\|R_k - R^*\| \geq \frac{\sigma(y)(R^*Q_k - P_k)(y)}{Q_k(y)}$$

$$\geq \sigma(y)(R^*Q_k - P_k)(y)$$

$$\geq c\,\|R^*Q_k - P_k\|$$

$$\geq c\,\varepsilon\,\|R^* - R_k\|$$

Hemos llegado a una contradicción, puesto que $R_k \neq R^*$ y $\gamma(R_k) \to 0$.

COROLARIO

Sea $R^* \equiv P^*/Q^*$ la mejor aproximación a f de la clase $R_m^n[a,b]$. Si $\min(n - \partial P^*, m - \partial Q^*) = 0$, entonces existe un $\gamma > 0$ tal que para todo $R \in R$,

$$||f - R|| \geq ||f - R^*|| + \gamma ||R - R^*||$$

Como en la teoría lineal, somos guiados a introducir en esta coyuntura, un operador de mejor aproximación, es decir, un operador $\mathbf{J}$ la cual selecciona de una clase R de funciones racionales generalizadas el elemento de la mejor aproximación a una función dada. Puesto que la existencia y - la unicidad de mejores aproximaciones estan aseguradas unicamente por hipótesis bastantes rígidas, nosotros adoptamos una aproximación ligeramente diferente. Dado R, nosotros definimos $\mathbf{J}$ de la siguiente forma. Para todo f, $\mathbf{J}f$ es el conjunto de todas las mejores aproximaciones a f en la clase R

$$\mathbf{J}f = \{R \in R / ||f - R|| = \min\}$$

Este conjunto puede ser vacío. En el caso de aproximación racional ordinaria, $R = R_m^n[a,b]$, $\mathbf{J}f$ nunca es vacío, y en verdad siempre contiene exactamente un elemento. Sin embargo aún en este caso ideal el operador $\mathbf{J}$ es discontinúo en algunos puntos de $C[a,b]$. Este descubrimiento se debe a Maehly y Witzgall.

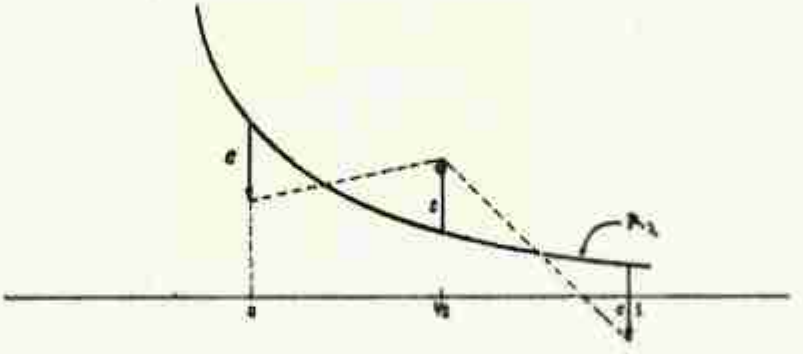
Permítanos hacer una pausa acá para dar un ejemplo de este fenómeno. Uno puede comenzar con las funciones racionales

$$R_\lambda(x) = \frac{\lambda}{\lambda + x} \quad (\lambda > 0) \qquad R_0(x) = 0$$

consideradas como elementos de $R1^{\circ}[0,1]$. Estos tienen la propiedad que

$$||R_\lambda - R_0|| = 1$$

para $\lambda > 0$, puesto que $R_\lambda(0) = 1$. Ahora determinemos una función contínua f_λ para la cual R_λ es la mejor aproximación en la clase considerada. - Por el teorema de alternación será suficiente dar $f_\lambda - R_\lambda$ tres puntos - de alternación. Por ejemplo, sea f_λ definida por la línea punteada en el



dibujo, siendo c cualquier constante en el intervalo (1/2, 1). Puesto que $\lambda \to 0$, f_λ converge uniformemente a una función contínua f_0 cuya mejor - aproximación es $R_0 \equiv 0$. Pero como hemos notado, R_λ no converge a R_0. Por lo tanto el operador de mejor aproximación $\mathfrak{J}$ es discontinuo en f_0. Obser- ve que para que R_0 sea la mejor aproximación de f_0, solamente dos puntos de alternación son requeridos.

Para las aproximaciones racionales generalizadas, incluyendo el caso racional ordinario, la continuidad de $\mathfrak{J}$ es gobernada por el siguiente teo- rema.

TEOREMA DE CONTINUIDAD

Si $R_o \in \mathbf{J}f_o$ y si $n(P + R_oQ) = \delta(P) + \delta(Q) - 1$, entonces $\mathbf{J}f$ es no vacío para toda f en un vecindario de f_o, y $\mathbf{J}$ es "contínuo" en f_o: hay un $\beta > 0$ tal que $||R_o - R|| < \beta||f_o - f||$ siempre que $R \in \mathbf{J}f$.

PRUEBA

La búsqueda para una mejor aproximación a f claramente puede ser confinada a aquellos $R \in \mathbf{R}$ para los cuales $||R - f|| \leq ||R_o - f||$. Por el teorema fuerte de unicidad tal R satisface las desigualdades.

$$\begin{aligned}\gamma||R - R_o|| &\leq ||f_o - R|| - ||f_o - R_o|| \\ &\leq ||f_o - f|| + ||f - R|| - ||f_o - R_o|| \\ &\leq ||f_o - f|| + ||f - R_o|| - ||f_o - R_o|| \\ &\leq ||f_o - f|| + ||f - f_o||\end{aligned}$$

Por consiguiente una mejor aproximación de f, si es que existe, debe satisfacer $||R - R_o|| \leq 2\gamma^{-1}||f - f_o||$. Asi β en el teorema puede ser tomado igual a $2\gamma^{-1}$. Queda por probarse que $\mathbf{J}f$ es no vacío.

Escriba $R_o = P_o/Q_o$, y asuma que $||P_o|| + ||Q_o|| = 1$. El número $2\varepsilon 1 \equiv \inf Q_o(x)$ es positivo. Ahora seleccione $\varepsilon_2 > 0$ tal que

$$\left.\begin{aligned}&||P|| + ||Q|| = 1 \\ &R = P/Q \in \mathbf{R} \\ &||R - R_o|| < \varepsilon_2\end{aligned}\right\} \Rightarrow ||Q - Q_o|| < \varepsilon_1$$

Para ver que esto es posible, suponga por el contrario que existe una sucesión $R_k = P_k/Q_k \in \mathbf{R}$ tal que $||P_k|| + ||Q_k|| = 1$, $R_k \to R_o$, y

$||Q_k - Q_o|| \geq \varepsilon_1$. Por la compactación podemos asumir que $P_k \to P$ y $Q_k \to Q$. Ya que $R_k \to R_o$, $P = R_oQ$. Por el lema 2, $P = P_o$ y $Q = Q_o$, una contradicción.

Ahora para completar la prueba, supongamos que $||f - f_o|| < \frac{1}{2}\gamma\varepsilon 2$. Entonces la mejor aproximación R de f(si es que existe) debe satisfacer - $||R - R_o|| < \varepsilon_2$. Si normalizamos $R = P/Q$ poniendo $||P|| + ||Q|| = 1$, entonces resultará que $||Q - Q_o|| < \varepsilon_1$. Ya que $Q_o(x) > 2\varepsilon 1$, $Q(x) \geq \varepsilon_1$. Así nuestra búsqueda para R es confinada a

$$\{P/Q \;/\; P\varepsilon P,\ Q \varepsilon Q,\ ||P|| + ||Q|| = 1,\ Q(x) \geq \varepsilon_1 \text{ en } [a,b]\}$$

Es elemental probar que este es un conjunto compacto, y debe contener por lo tanto una mejor aproximación a f.

4. ALGORITMOS

Consideremos nuevamente el problema de aproximar una función dada, - $f \in C[a,b]$ mediante una función racional generalizada de la forma

$$R = \frac{P}{Q} = \frac{a_o g_o + \ldots + a_n g_n}{b_o h_o + \ldots + b_m h_m}$$

en la cual $Q(x) > 0$ en $[a,b]$. La clase de todos los R tales, obtenidos mediante la variación de los parámetros a_i y b_i, se denota como es usual- por R

Uno de los algoritmos más simples puede estar basado en el uso de - sistemas de desigualdades lineales. Supóngase que para un cierto valor positivo de ε nosotros preguntamos por un $R = P/Q$ que satisfaga $|f(x) - R(x)| \leq \varepsilon$. Nosotros podemos hacer arreglos que $Q(x) \geq 1$ en $[a,b]$ multiplicando el numerador y el denominador por un número positivo apropiado. La desigualdad $-\varepsilon \leq f(x) - R(x) \leq \varepsilon$ puede escribirse en la forma equivalente-

$$-\varepsilon Q(x) \leq f(x)\,Q(x) - P(x) \leq \varepsilon\, Q(x)$$

y por lo tanto las condiciones a ser impuestas sobre P y Q son justamente- las siguientes:

$$(1) \qquad \left\{ \begin{array}{c} -Q(x) \leq -1 \\ f(x)\,Q(x) - P(x) \leq \varepsilon\, Q(x) \\ -f(x)\,Q(x) + P(x) \leq \varepsilon\, Q(x) \end{array} \right\} \quad (a \leq x \leq b)$$

Este sistema de desigualdades lineales en el vector coeficiente

$$[a_o, \ldots, a_n, b_o, \ldots, b_m]$$

es consistente o inconsistente. En el primer caso cualquier solución del

sistema proporciona la aproximación buscada. En el último caso, ε es tan-pequeño que no existe elemento de R dentro de una distancia ε desde f. Si-se desea una mejor aproximación, entonces el valor de ε debe ser ajustado-hasta que se encuentre un ε mínimo para el cual el sistema (1) sea consisten-te. La prueba para consistencia de (1) puede hacerse buscando el mínimo de-la función convexa:

$$\delta = \max_{x} \max \{1 - Q,\ fQ - P - \varepsilon Q,\ P - fQ - \varepsilon Q\}$$

Aquí δ es una función del vector coeficiente $c = [a_0,\ldots,a_n,b_0,\ldots b_m]$ mientras que P, f y Q son funciones de x en el rango $[a,b]$. El método del capítulo 1 puede utilizarse para encontrar ese valor de c para el cual $-\delta(c)$ es un mínimo. Si $\delta(c) \leq 0$, entonces c es una solución del sistema (1). Si $\delta(c) > 0$, entonces (1) es inconsistente y ε se escogió demasiado pequeño. Este algoritmo es conocido como el "Método de la Desigualdad Lineal".

Otro procedimiento en el cual es facilmente explicado puede ser llama-do el "Algoritmo Minimax de Peso". La función cuyo mínimo buscamos se es--cribe en la forma

$$\Delta = \max_{a\leq x\leq b} \left| f(x) - \frac{P(x)}{Q(x)} \right| = \max_{a\leq x\leq b} \frac{1}{Q(x)} \left| f(x)Q(x) - P(x) \right|$$

la cual inmediatamente sugiere aplicar el método de iteración. Por lo tan-to nosotros introducimos índices en la expresión que aparece arriba y mi-nimizamos a cambio la función

$$\delta_k = \max_{a\leq x\leq b} \frac{1}{Q_{k-1}(x)} \left| f(x)Q_k(x) - P_k(x) \right|$$

En el k-ésimo paso del proceso $1/Q_{k-1}(x)$ se mantiene fijo en el valor determinado del paso precedente mientras que Q_k y P_k son variados para hacer δ_k un mínimo. A fin de evitar una solución trivial, uno de los coeficientes en P ó Q se hace igual a una constante. La minimización de δ se realiza mediante los métodos del capítulo 1. Aún cuando este método se programa facilmente y trabaja muy bien en la práctica, hace falta para ello un teorema de convergencia.

El próximo algorítmo a ser considerado es similar en espiritu pero ligeramente más dificil a realizar, sin embargo posee un teorema de convergencia. En el paso k-ésimo de este algorítmo una aproximación de $R_k = P_k/Q_k$ será disponible de el paso precedente. Computamos entonces el número

$$\Delta_k = ||f - R_k||$$

Ahora definimos una función auxiliar

$$\delta_k(R) = \max_x \{|f(x)Q(x) - P(x)| - \Delta_k Q(x)\}$$

Seleccionemos $R_{k+1} = P_{k+1}/Q_{k+1}$ como para minimizar δ_k bajo la restricción que $||Q_{k+1}|| = 1$. Es claro que si la $||P||$ es grande, entonces $\delta_k(R)$ es grande, asi que en la minimización de δ_k no sea necesario restringir $||P_{k+1}||$. Al principio, R_o puede ser arbitrario excepto que su denominador debería ser positivo en $[a,b]$. Si $\delta_k(R_{k+1}) \geq 0$, entonces nosotros nos detenemos y R_k es una mejor aproximación de f. Este procedimiento es conocido como el "Algorítmo de Corrección Diferencial". Probamos ahora que ello es efectivo.

TEOREMA

En el algorítmo de corrección diferencial, $\Delta_k \downarrow \Delta^* = \inf \Delta$. Si existe una mejor aproximación, entonces la convergencia es al menos lineal: $\Delta_{k+1} - \Delta^* \leq \theta(\Delta_k - \Delta^*)$ con $\theta < 1$.

PRUEBA

Si el denominador Q, de R es positivo en todo $[a,b]$, entonces $\delta_k(R)$ puede ser escrito en la forma siguiente:

$$\delta_k(R) = \max_x \{[|f(x) - R(x)| - \Delta_k] Q(x)\} \tag{2}$$

Ahora supongamos que existe un índice k tal que $\inf_{a<x<b} Q_{k+1}(x) \leq 0$. Podemos tomar k como el primero de tal índice. Puesto que $Q_0(x) > 0$, $k \geq 0$. También $R_k \in R$.

Probaremos que R_k es una mejor aproximación a f. En el caso contrario existe $R = P/Q \in R$ tal que $||Q|| = 1$ y $\Delta(R) < \Delta(R_k)$. Por lo tanto

$$|f(x) - R(x)| < \Delta_k$$

para todo x, y consecuentemente de (2), $\delta_k(R_{k+1}) \leq \delta_k(R) < 0$. Pero si $Q_{k+1}(x_o) \leq 0$, nosotros obtenemos la desigualdad contradictoria

$$\delta_k(R_{k+1}) \geq |f(x_o) Q_{k+1}(x_o) - P(x_o)| - \Delta_k Q_{k+1}(x_o) \geq 0.$$

Por lo tanto a menos que el algorítmo produzca una solución en un número finito de pasos, nosotros podemos asumir que $Q_k(x) > 0$ para todo k y x. Ahora nosotros podemos probar que $\delta_k(R_{k+1}) \leq 0$, la igualdad ocurre unicamente si R_k es una mejor aproximación. En realidad ocurre unicamente si R_k es una mejor aproximación. En realidad $\delta_k(R_{k+1}) \leq \delta_k(R_k) = 0$ de -

(2), mientras que si R_k no es una mejor aproximación, entonces como anteriormente nosotros podemos mostrar que $\delta_k(R_{k+1}) < 0$. Luego nosotros establecemos que $\Delta_0 > \Delta 1 > \ldots$. En realidad,

$$0 > \delta_k(R_{k+1}) \geq \max_x \{|f(x) - R_{k+1}(x)| - \Delta_k\} = \Delta_{k+1} - \Delta_k,$$

donde nosotros hemos usado el hecho que $||Q_{k+1}|| = 1$ en la ecuación (2). Entonces la sucesión $\{\Delta_k\}$ converge en dirección hacia abajo en un límite-L. Sea $\Delta^* = \inf_{R \in R} \Delta(R)$. Queda a ser mostrado que $\Delta^* = L$. Si no es asi, - entonces existe un $R = P/Q \in R$ tal que $\Delta(R) < L \leq \Delta_k$, de donde $\delta_k(R_{k+1}) \leq \delta_k(R) \leq \alpha \max_x \{|f(x) - R(x)| - \Delta_k\} = \alpha[\Delta(R) - \Delta_k] + \Delta_k$. Ahora permitiendo que $k \to \infty$, nosotros obtenemos $L \leq \alpha [\Delta(R) - L] + L$, lo cual es - una contradicción. La aseveración acerca de la convergencia lineal se prueba de la siguiente manera. Sea R una mejor aproximación a f. Entonces mediante el argumento precedente nosotros tenemos $\Delta_{k+1} - \Delta_k \leq \alpha(\Delta^* - \Delta_k)$. Por lo tanto $\Delta_{k+1} - \Delta^* = (\Delta_k - \Delta^*) + (\Delta_{k+1} - \Delta_k) \leq \Delta_k - \Delta^* + \alpha(\Delta^* - \Delta_k) = (1 - \alpha)(\Delta_k - \Delta^*)$. Luego $||Q|| = 1$, $0 \leq 1 - \alpha < 1$.

COROLARIO

Sea R^k una mejor aproximación en R a f. Si $\eta(p + R^k Q) = n + m + 1$, entonces las funciones racionales producidas por el algoritmo de corrección diferencial convergen al menos linealmente en R^k:

$$||R_k - R^*|| \leq A\theta^k \qquad (\theta < 1)$$

PRUEBA

Mediante el teorema fuerte de unicidad (sec. 3),

$$||R_k - R^{*}|| \leq \gamma^{-1}\left[||f - R_k|| - ||f - R^*||\right] = \gamma^{-1}(\Delta_k - \Delta^*)$$

Mediante el teorema precedente esto es mayorizado por

$$\gamma^{-1}(1 - \alpha)^k(\Delta_o - \Delta^*)$$

En la aplicación práctica del algoritmo de corrección diferencial, - la restricción de $||Q|| = 1$ es menos conveniente que $|b_i| \leq 1$, donde

$$Q = \sum_{i=0}^{m} b_i h_i.$$

La prueba de convergencia se mantiene válida si nosotros escribimos

$$0 > \delta_k(R_{k+1}) \geq \beta(\Delta_{k+1} - \Delta_k), \text{ con } \beta = \max_{|b_i|\leq 1} ||Q|| .$$

La constante $1 - \alpha$ es reemplazada por $1 - \alpha\beta^{-1}$ al final y esto tambiénes un número en el intervalo $[\,0,1\,]$. La minimización en δ_K es un problema- de "Programación Convexa", o su extensión al caso de una gran cantidad de funciones γ_i.

También se conoce un algorítmo en el cual las aproximaciones racionales sucesivas convergen cuadráticamente (bajo ciertas condiciones) a la - mejor aproximación. El lector interesado puede tomar como referencia el - análisis de H. Werner publicado en una serie de artículos [1962, a, 1963]

APENDICE

TEOREMA DE EXISTENCIA (de las mejores Aproximaciones) 1.

Un sub-espacio lineal dimensional finito de un espacio lineal normado contiene por lo menos un punto de distancia mínima desde un punto fijo.

PROPIEDAD

La función Δ es convexo

$$\Delta(c) = \Delta(c_1,\ldots,c_n) = \left|\left| \sum_{i=1}^{n} c_i f_i - g \right|\right|$$

PRUEBA de que Δ es convexo.

$$\begin{aligned}
\Delta(\lambda c + \mu d) &= ||\Sigma(\lambda c_i + \mu d_i)f_i - g|| \\
&= ||\lambda(\Sigma c_i f_i - g) + \mu(\Sigma d_i f_i - g)|| \\
&\leq \lambda||\Sigma c_i f_i - g|| + \mu||\Sigma d_i f_i - g|| \\
&= \lambda\Delta(c) + \mu\Delta(d)
\end{aligned}$$

TEOREMA 2.

Un mínimo local de una función convexa es necesariamente un mínimo global.

OBSERVACION

Correspondiente a cada conjunto A en un espacio lineal hay un conjunto convexo $\mathcal{K}(A)$ llamado su cápsula convexa, el cual se define como-

el conjunto de los puntos g los cuales son expresables como sumas finitas de la forma $g = \Sigma\theta_i f_i$ con $f_i \in A$, $\theta_i \geq 0$ y $\Sigma\theta_i = 1$. Tales combinaciones lineales son convenientemente llamadas combinaciones lineales convexas.

TEOREMA DE LAS DESIGUALDADES LINEALES 3.

Sea U un subconjunto compacto de $\mathbb{R}_n$. Una condición suficiente y necesaria para que el sistema de desigualdades lineales

$$\langle u,z\rangle > 0 \quad (u \in U)$$

sea inconsistente es que $0 \in \mathcal{H}(U)$.

TEOREMA DE CARATHEODORY 4.

Sea A un subconjunto de un espacio lineal n-dimensional. Cada punto de la cápsula convexa de A puede ser expresado como una combinación lineal convexa de n + 1 (o menos) elementos de A.

TEOREMA DE GRAM-SCHMIDT 5.

Sea {f1, f2,...} un conjunto de vectores linealmente independiente en un espacio con producto interno. Para cada n es posible definir un vector g_n como una combinación lineal de $f1,\ldots,f_n$ en tal forma que $\{g_1, g_2, \ldots\}$ es ortonormal.

TEOREMA 6.

En un espacio con producto interno, si un conjunto $\{g1,\ldots,g_n\}$ es independiente entonces la matriz (Gram) que tiene los elementos $A_{ij} = \langle g_i, g_j \rangle$ es no singular.

PRUEBA

Si el proceso Gram-Schmidt (T.5 de Apéndice) se aplica a $\{g1,\ldots,g_n\}$ obtenemos un conjunto ortonormal $\{f_1,\ldots,f_n\}$ en el cual cada f_i es de la forma $f_i = \sum_{j=1}^{n} B_{ij}g_j$ (En realidad $B_{ij} = 0$ si $j > i$). Asi

$$\delta_{ij} = \langle f_i, f_j \rangle = \langle \Sigma_\nu B_{i\nu}g_\nu, \Sigma_\mu B_{j\mu}g_\mu \rangle = \Sigma_\nu \Sigma_\mu B_{i\nu} \langle g_\nu, g_\mu \rangle B_{j\mu}$$

Esta última ecuación puede ser escrita en la forma de matriz $I = BAB^T$ de la cual se deduce que A es no singular.

TEOREMA 7.

Denotemos con $\{g_1,\ldots,g_n\}$ un conjunto ortonormal en un espacio - con producto interno con la norma definida por $||h|| = \sqrt{\langle h,h \rangle}$.

La expresión $||\Sigma c_i g_i - f||$ es un mínimo si y sólo si $c_i = \langle f,g_i \rangle$

TEOREMA 8.

La prueba M de Weierstrass. Si $|f_n(x)| \leq M_n$ y $\Sigma M_n < \infty$, entonces - Σf_n converge uniformemente.

TEOREMA 9.

El espacio C [a,b] es completo.

TEOREMA 10.

En cualquier espacio con producto interno, definida

$$||f|| = + \sqrt{\langle f,f \rangle}$$

Entonces los siguientes son verdaderos:

i) $|\langle f,g\rangle| \leq ||f||\,||g||$ (desigualdad de Cauchy Schwarz).

ii) $||f + g|| \leq ||f|| + ||g||$ (desigualdad triangular).

iii) $||f + g||2 + ||f - g||2 = 2||f||2 + 2||g||^2$ (Ley del paralelogramo).

iv) $\langle f,g\rangle = 0 \Rightarrow ||f - g||2 = ||f||2 + ||g||^2$ (Ley pitagoriana).

TEOREMA DE LA VALLEE POUSSIN 11.

Si P es un polinomio generalizado tal que f - P asume valores alternadamente positivos y negativos en n + 1 puntos consecutivos x_i de $[a,b]$, entonces $E(f) \geq \min_i |f(x_i) - P(x_i)|$

TEOREMA DE ALTERNACION 12.

Sea $\{g_1,\ldots,g_n\}$ un sistema de elementos de $C[a,b]$ satisfaciendo la condición de Haar, y sea X cualquier subconjunto cerrado de $[a,b]$. Para que un cierto polinomio generalizado $P - \Sigma c_i g_i$ sea una mejor aproximación en X para un $f \in C[X]$ dado es necesario y suficiente que la función de error r = f - P exhiba en X por lo menos n + 1 "alternaciones" - asi: $r(x_i) = -r(x_{i-1}) = \pm ||r||$, con $x_o < \ldots < x_n$ y $x_i \in X$. Aqui

$$||r|| = \max_{x \in X} |r(x)|$$

TEOREMA DE CARACTERIZACION 13.

Para que los coeficientes $c_1,\ldots,c_n$ puedan representar la norma - (uniforme) de $r = \Sigma c_i g_i - f$ un mínimo, es necesario y suficiente que el origen del espacio n-dimensional esté situado en la cápsula convexa del-

conjunto del punto $\{r(x)\hat{x} \;/\; |r(x)| = ||x||\}$, donde $\hat{x}$ denota la n-upla $[g_1(x),\ldots,g_n(x)]$

LEMA 1

Sea $\{g_1,\ldots,g_n\}$ un sistema de elementos de $C[a,b]$ satisfaciendo la condición de Haar. Sea $a \leq x_o < x_1 <\ldots< x_n \leq b$, y sean $\lambda_o,\ldots,\lambda_n$ constantes no ceros. Para que 0 esté situado en la cápsula convexa de las n-uplas $\lambda_o\hat{x}_o,\ldots, \lambda_n\hat{x}_n$ es necesario y suficiente que los λ alternen en signos $\lambda_i\lambda_{i-1} < 0$ para $i = 1, \ldots, n$

BIBLIOGRAFIA

1.- INTRODUCTION TO APPROXIMATION THEORY
E. W. CHENEY
Mc Graw-Hill Book Company, 1976

2.- OLDS. C. D. (1950) THE BEST POLINOMIAL APPROXIMATION OF FUNCTIONS
American Mathematical Monthly 57, 617 - 621.

3.- CLENDENIN, W. W. (1961) NOTES ON THE CONSTRUCTION OF RATIONAL APPROXIMATIONS FOR THE ERROR FUNCTION AND FOR SIMILAR FUNCTIONS, ACM COMMUNICATIONS 4, 354 - 355.