UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE CIENCIAS

LA INVERSA GENERALIZADA

T E S I S

QUE PARA OBTENER EL TITULO DE

M A T E M A T I C O

P R E S E N T A

JESUS LOPEZ ESTRADA

MEXICO, D. F. 1973

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Con todo cariño a mis padres como justo tributo a su abnegación y desvelos.

Con todo cariño a mi esposa e hijos compañeros de ilusiones y anhelos, porquienes siempre lucharé para alcanzar el ideal que juntos forjamos.

Con todo cariño a mis hermanos.

A mis amigos y compañeros.

EXAMENES
PROFESIONALES

INDICE

PROLOGO

La comunicación de las ideas siempre ha sido un problema para el hombre, y lo es más marcado en ciertas disciplinas del saber humano, como ocurre en las Matemáticas.

Creemos que para lograr una presentación accesible de cierto tema de las Matemáticas a un gran número de personas es aconsejable partir de una serie de problemas concretos con un mismo común denominador y sin condicionamientos a priori, en términos de los cuales, se introduzcan y se desarolle la teoría, necesarios para su solución.

En términos de los lineamientos anteriores nosotros desarrollamos el presente trabajo, lo que creemos haber logrado, si no en su totalidad, sí en una gran parte.

Nosotros empezamos con una reseña histórica de la Inversa Generalizada, tema al cual está dedicada la presente tesis.

En el capítulo I, discutimos dos problemas del campo de la Estadística en términos de los cuales introducimos los conceptos de la mejor solución aproximada y de la inversa generalizada como "generalizaciones" de solución de un sistema lineal de ecuaciones algebráicas y de la inversa de una matriz, respectivamente.

En el capítulo II, estudiamos la inversa generalizada de una matriz via la descomposición singular, la caracterizamos y damos una aplicación al Análisis Numérico.

Por último en el capítulo III, generalizamos los conceptos introducidos en el capítulo I a ecuaciones dadas por transformaciones lineales continuas entre espacios de Hilbert, y damos importantes aplicaciones a las ecuaciones integrales de Fredholm.

Agradecemos la valiosa orientación y ayuda prestada por el maestro y el amigo Dr. Pablo Barrera para lograr la realización del presente trabajo.

Finalmente, agradecemos al CIMAS* el beneficio otorgado al autor para el desarrollo del presente trabajo.

---

* Centro de Investigación en Matemáticas Aplicadas y en Sistemas.

RESEÑA HISTORICA
DE LA
INVERSA GENERALIZADA

En 1920, E.H. Moore introduce en [7] por primera vez la Inversa Generalizada de una matriz. Y en 1935, presenta en [8] un análisis general de la inversa generalizada para operadores lineales.

En 1955, R. Penrose introduce en [9] independientemente de E.H. Moore la Inversa Generalizada de una matriz. Y un año después, en [10] hace notar que la solución obtenida mediante la aplicación de la matriz Inversa Generalizada a un sistema de ecuaciones lineales algebraicas es su solución aproximada según el criterio de los mínimos cuadrados, la aplica a problemas de Estadística y da dos métodos para su cálculo.

Los trabajos de Penrose estimulan el estudio y la aparición de una amplia producción literaria sobre la Inversa Generalizada con variadas aplicaciones y generalizaciones a transformaciones lineales entre ciertos espacios lineales normados.

Pyle en [12] aplica en su método del gradiente a la

Inversa Generalizada para resolver problemas de programación lineal. Rosen en [13 y 14] siguiendo a Pyle utiliza la Inversa Generalizada en su método del gradiente para resolver problemas de programación lineal y no-lineal.

Kalman en [15 y 16] y Florentin en [17] aplican la Inversa Generalizada a problemas de control óptimo en teoría del control.

Loud en [18 y 19] entre otros introduce la Inversa Generalizada a funciones y matrices de Green y da un método para calcularlas.

Petryshyn en [23] hace un estudio detallado de la Inversa Generalizada para transformaciones lineales continuas entre espacios de Hilbert y da un método para calcularla. Otra caracterización importante de la Inversa Generalizada para el caso del presente párrafo la da Showalter en [22].

Kammerer y Nashed en [21] entre otros aplican la Inversa Generalizada para transformaciones lineales continuas entre espacios de Hilbert a las ecuaciones integrales de Fredholm inconsistentes de primera y segunda clase y dan un método iterativo para calcularla.

En [11] se encuentran notas históricas con más detalle de la Inversa Generalizada con una amplia bibliografía.

CAPITULO I

INTRODUCCION

En este capítulo introductorio ilustramos con dos ejemplos las ideas centrales que nos permitirán asociarle a una transformación lineal $A:R^n \rightarrow R^m$ una única transformación lineal $B:R^m \rightarrow R^n$ que sea la inversa de A cuando ésta existe o una generalización de la inversa de A cuando ésta no existe.

1. EL CRITERIO DE LOS MINIMOS CUADRADOS.

(a) Planteamiento.

Un problema que se plantea con frecuencia consiste en "ajustar" una línea recta a un número finito de puntos en el plano obtenidos experimentalmente.

Por ejemplo supóngase que los siguientes puntos

$$p_1 = (t_1, v_1),\ p_2 = (t_2, v_2), \ldots, p_n = (t_n, v_n)$$

son obtenidos experimentalmente, en donde la v_k denotan el volumen de cierto gas que a presión constante es sometido a una temperatura t_k. Ahora el problema es encontrar los valores de m y b tales que la línea recta $v=mt+b$ se "ajuste" en "lo mejor posible" a los puntos $p_1, p_2, \ldots, p_n$. Véase la figura 1 abajo.

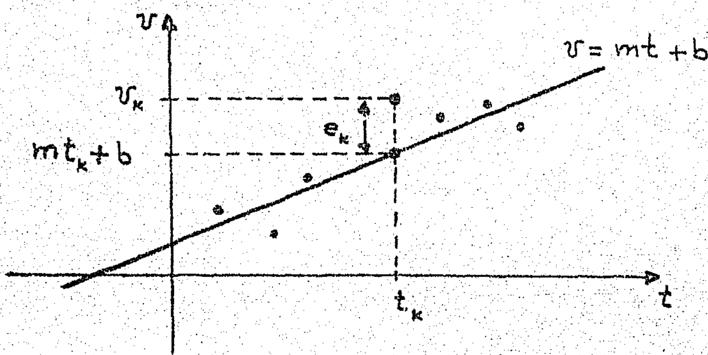

Fig. 1. Ilustración del problema de ajuste de una línea recta a un conjunto de puntos dados en el plano.

Un criterio muy usado para resolver este problema es el

<u>Criterio de los Mínimos Cuadrados</u>, el cual consiste en calcular la ecuación de la línea recta $v=mt+b$ tal que

$$(1) \qquad \sum_{k=1}^{n} e_k{}^2$$

sea mínima, en donde

$$(2) \qquad v_k = mt_k+b+e_k.$$

como se ilustra en la figura 1.

Ahora el sistema de ecuaciones (2) puede escribirse como

$$\bar{c} = A\bar{x} + \bar{e}, \tag{3}$$

en donde

$$A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix}, \quad \bar{c} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} m \\ b \end{bmatrix}, \text{ y } \quad \bar{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}. \tag{4}$$

Luego el criterio de los mínimos cuadrados se traduce en encontrar un vector $\bar{X}_0^{\,t} = (m_0, b_0)$ que minimice a

$$\sum_{k=1}^{n} e_k^{\,2} = ||\bar{e}||^2 = ||A\bar{x}-\bar{c}||^2. \tag{5}$$

Obsérvese que la matriz dada en (4) está formada por vectores columna linealmente independientes.

(b). <u>Generalización.</u>

En general el problema clásico de los mínimos cuadrados consiste en encontrar un vector $\bar{x}_0$ en R^n tal que

$$||A\bar{x}_0 - \bar{c}||^2 \leq ||A\bar{x}-\bar{c}||^2, \tag{6}$$

para todo vector $\bar{x}$ en R^n, en donde A es una matriz real dada de dimensiones mxn, cuyos vectores columna son linealmente independientes (l.i.), y $\bar{c}$ es un vector en R^m también dado.

Si consideramos a A como una transformación lineal $A:R^n \rightarrow R^m$, entonces geométricamente el criterio de los mínimos cuadrados consiste en calcular la imagen inversa de la proyección "ortogonal" $\bar{c}_A$ de $\bar{c}$ sobre el subespacio imagen de A (que denotaremos por Im(A)) sumergido en R^m. Véase la figura 2 abajo.

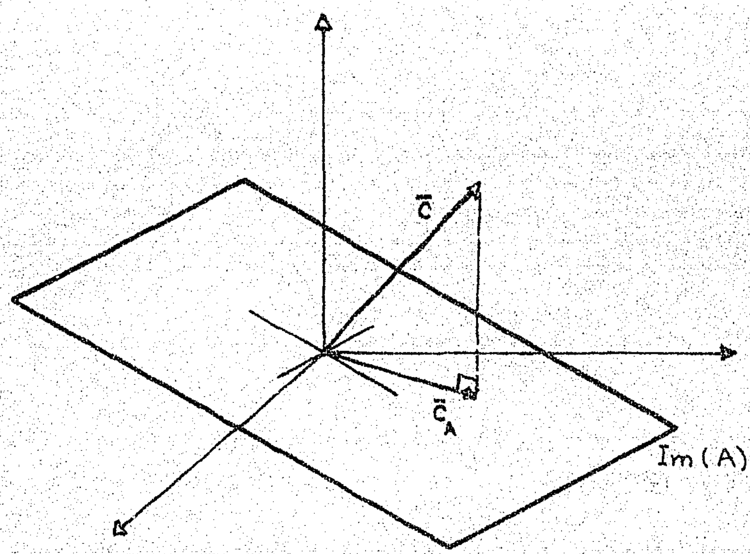

Figura 2. Interpretación geométrica del problema clásico de los mínimos cuadrados para el caso n=2 y m=3.

De la discusión del párrafo anterior y del hecho que A es una matriz de dimensiones mxn y con n vectores columna l.i. se sigue que el problema clásico de los mínimos cuadrados siempre tiene una única solución $\bar{x}_0$.

Ahora aplicando las técnicas del cálculo de máximos y mínimos del cálculo diferencial para calcular $\bar{x}_0$

nosotros tenemos que

$$
\begin{aligned}
0 &= \nabla(||\bar{c}-A\bar{x}||^2)|_{\bar{x}=\bar{x}_0} \\
&= \nabla((\bar{c}-A\bar{x})^t\ (\bar{c}-A\bar{x}))|_{\bar{x}=\bar{x}_0} \\
&= 2(A^tA\bar{x}_0-A^t\bar{c}).
\end{aligned}
$$

Esto es, el punto $\bar{x}_0$ que minimiza a $||\bar{c}-A\bar{x}||^2$ es una solución del sistema

$$(7) \qquad A^tA\ \bar{x} = A^t\bar{c}\ ,$$

conocida por la ecuación normal asociada al sistema lineal inconsistente

$$(8) \qquad A\bar{x} = \bar{c}.$$

Ahora como A^tA es de orden nxn y el rango de A^tA es igual al rango de A que es igual a n, tenemos que el punto

$$\bar{x}_0 = (A^tA)^{-1}\ A^t\ \bar{c}$$

satisface a (6).

La discusión anterior se resume en el siguiente

Teorema 1. La solución del problema clásico de los mínimos cuadrados está dada por

(9) $$\overline{x}_0 = (A^tA)^{-1} A^t\overline{c} .$$

(c) Una Aplicación a Sistemas de Ecuaciones Lineales Algebraicas.

Considérese el sistema lineal

(10) $$A \overline{x} = \overline{c}$$

en donde A es una matriz real de dimensiones mxn con m>n y de rango n. Y supóngase que (10) no tiene una solución. Entonces por el teorema 1 podemos asociarle a (10) una solución aproximada $\overline{x}_0$ según el criterio de los mínimos cuadrados, la cual es única y está dada por

$$\overline{x}_0 = (A^tA)^{-1} A^t \overline{c} .$$

Por último obsérvese que si (10) tiene solución ésta coincide con su solución aproximada.

2. ESTIMACION LINEAL PUNTUAL

(a) Planteamiento y Discusión.

En esta sección vamos a discutir con un ejemplo de estimación lineal puntual el problema dual expuesto en la sección anterior.

El modelo general de un problema clásico de regresión lineal consiste en calcular una solución del sistema

$$(1) \qquad \hat{b} = Ax + \hat{\varepsilon},$$

en donde A es una matriz real dada de dimensiones mxn con m>n y de rango n, $\hat{b}$ y $\hat{\varepsilon}$ son vectores en R^m obtenidos experimentalmente, y $\bar{x}$ es un vector en R^n de parámetros desconocidos que son los que se quieren estimar. El vector $\hat{\varepsilon}$ es un vector "estocástico" que nos da los errores cometidos en las mediciones efectuadas el cual generalmente satisface

$$(2) \qquad E(\hat{\varepsilon}) = \bar{0}$$

y

$$(3) \qquad E(\hat{\varepsilon}\hat{\varepsilon}^t) = S,$$

con S una matriz real de dimensiones mxm positiva definida.

El vector $E(\hat{\varepsilon})$ denota el vector error esperado, y $E(\hat{\varepsilon}\hat{\varepsilon}^t)$ denota la matriz de "covarianza" de las "variables aleatorias" $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ que forman el vector $\hat{\varepsilon}$.

Otra descripción equivalente del modelo clásico de regresión lineal nos la dan las siguientes ecuaciones

$$(4) \qquad E(\hat{b}) = A\bar{x}$$

y

$$(5) \qquad E((\hat{b}-A\overline{x})(\hat{b}-A\overline{x})^t) = S,$$

con S una matriz positiva definida.

Teóricamente, la solución $\overline{x}_0$ de un problema clásico de regresión lineal es única y la podemos determinar si conocemos completamente una de las "distribuciones" de $\hat{b}$ o $\hat{\varepsilon}$. Pero en la práctica sólo conocemos los valores b_k observados, y por lo tanto, no podemos determinar a $\overline{x}_0$.

De la discusión del párrafo anterior, el camino natural a seguir es el de estimar a $\overline{x}_0$. Pero antes de estimar a $\overline{x}_0$ tratemos de estimar su k-ésima componente x_{k0} llevando el problema de estimación vectorial a un problema de estimación escalar, el cual es natural pensar que es más sencillo.

El problema de estimar la k-ésima componente x_{k_0} de $\overline{x}_0$, se puede formular de la siguiente manera:

Estimar el valor en $\overline{x}_0$ de la funcional

$$(6) \qquad \phi(\overline{x}) = \overline{c}^t\overline{x} ,$$

en donde $\overline{c}$ es un vector dado en R^n. Obsérvese que en particular, si $\overline{c}=\overline{e}_k$ en donde $\overline{e}_k$ es el vector cuya k-ésima componente es la unidad y sus demás componentes son cero entonces se estima a x_{k0}.

Para nuestros propósitos consideraremos la funcional estimadora

$$(7) \qquad \hat{\phi} = \bar{u}^t \hat{b} \ ,$$

en términos de la cual nuestro problema es calcular el vector coeficiente $\bar{u}^t$ en R^m de manera que $\hat{\phi}$ nos dé una buena estimación de ϕ .

A un estimador $\hat{\phi}$ dado por (7) lo llamaremos <u>no-sesgado</u> si $E(\hat{\phi}) = \phi$. En términos de (6) y (7) si

$$(8) \qquad E(\bar{u}^t \hat{b}) = \bar{c}^t \, \bar{x} \ ,$$

para todos los valores de $\bar{x}$.

Ahora como

$$E(\bar{u}^t \hat{b}) = \bar{u}^t \, E(\hat{b}) = u^t A\bar{x},$$

tenemos que $\hat{\phi} = \bar{u}^t \hat{b}$ es un estimador no-sesgado de $\phi = \bar{c}^t \, \bar{x}$ si $\bar{u}$ satisface la ecuación

$$\bar{u}^t \, A\bar{x} = \bar{c}^t \, \bar{x}$$

para todos los valores de $\bar{x}$. Esto es, $\hat{\phi} = \bar{u}^t \, \hat{b}$ es un estimador no sesgado de $\phi = \bar{c}^t \, \bar{x}$ si $\bar{u}$ es una solución de la ecuación

$$(9) \qquad A^t \bar{u} = \bar{c}.$$

Pero como m>n tenemos que (9) tiene una infinidad de soluciones $\bar{u}$, en efecto cualquier $\bar{u}$ del conjunto

$$\{\bar{u} \mid \bar{u}=\bar{u}_p+\bar{n}, \text{ con } A^t\bar{n}=0 \text{ y } A^t \bar{u}_p=\bar{c}\}$$

es una solución de (9). Lo que nos lleva al problema de dar un criterio que nos permita escoger sin ambiguedades el estimador no-sesgado <u>más eficiente</u> $\hat{\phi} = \bar{u}^t \hat{b}$ de $\phi = \bar{c}^t \bar{x}$.

Una propiedad deseable para un buen estimador no-sesgado es que su "varianza" sea <u>mínima</u>. Lo que nos proporciona un criterio para escoger el estimador no-sesgado y <u>más eficiente</u> de $\phi = \bar{c}^t \bar{x}$, el estimador $\hat{\phi} = \bar{u}^t \bar{b}$ de varianza mínima.

Calculando la varianza de $\hat{\phi} = \bar{u}^t \hat{b}$

$$\begin{aligned}
\sigma^2(\hat{\phi}) &= E((\hat{\phi}-E(\hat{\phi}))^2) \\
&= E((\bar{u}^t \hat{b} - \bar{u}^t A\bar{x})^2) \\
&= \bar{u}^t E((\hat{b}-A\bar{x}) (\hat{b}-A\bar{x})^t)\bar{u} \\
&= \bar{u}^t S \bar{u}
\end{aligned}$$

Nosotros tenemos la siguiente formulación equivalente del problema de determinar el estimador no-sesgado y más eficiente de $\phi = \bar{c}^t \bar{x}$:

Minimizar la funcional

(10) $$\psi(u) = \bar{u}^t \, S \, \bar{u}$$

con restricciones

(11) $$A^t \, \bar{u} = \bar{c} \, .$$

Con objeto de dar una interpretación geométrica a nuestro problema de estimación lineal denotemos por R^m_S al espacio vectorial R^m con producto interior

(12) $$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t \, S \, \bar{y}.$$

En estos términos el estimador no-sesgado y más eficiente $\hat{\phi} = \bar{u}^t \, \hat{b}$ de $\phi = \bar{c}^t \, \bar{x}$ está dado por $\hat{\phi}_0 = \bar{u}_0^t \, \hat{b}$, en donde $\bar{u}_0$ es la solución de menor norma en R^m_S de (11). Véase la figura 3 a continuación.

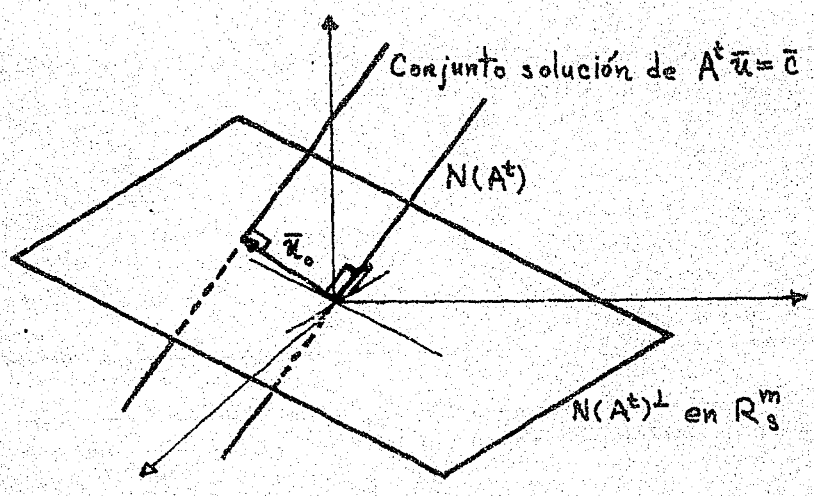

Figura 3. Interpretación Geométrica del Criterio de Esti-

mación Lineal de Mínima Varianza para el caso m=2 y n=3. Aquí $N(A^t)$ denota el conjunto solución de $A^t \bar{u} = 0$ y $N(A^t)^{\perp}$ el subespacio ortogonal de $N(A^t)$ en R^m_S.

Del párrafo anterior se tiene que el estimador no-sesgado de mínima varianza $\hat{\phi}_0 = \bar{u}_0^t \hat{b}$ de $\phi = \bar{c}^t \bar{x}$ es único.

Usando la técnica de los multiplicadores de Lagrange para calcular $\bar{u}_0$. Nosotros pasamos a minimizar la funcional

$$(13) \qquad L(\bar{u},\bar{\lambda}) = \bar{u}^t S \bar{u} - 2(A^t \bar{u}-\bar{c})^t \bar{\lambda},$$

en donde $\bar{\lambda}$ es un vector multiplicador de Lagrange en R^m.

Calculando $\frac{\partial L}{\partial u}, \frac{\partial L}{\partial \lambda}$ e igualando al vector cero, se tiene que

$$\bar{u}_0 = S^{-1} A\bar{\lambda}_0 \quad \text{y } A^t \bar{u}_0 = \bar{c}.$$

Esto es, $\bar{u}_0$ está dado por

$$(14) \qquad \bar{u}_0 = S^{-1} A\bar{\lambda}_0$$

en donde $\bar{\lambda}_0$ es una solución de la ecuación

$$(15) \qquad A^t S^{-1} A \bar{\lambda} = \bar{c}.$$

Pero como $A^t S^{-1} A$ es una matriz de orden nxn con rango igual al rango de A que es n, nosotros tenemos que $A^t S^{-1} A$ es invertible y por lo tanto $\bar{u}$ está dado por

$$\bar{u}_0 = S^{-1} A(A^t S^{-1} A)^{-1} \bar{c}.$$

La discusión anterior se resume en el siguiente

<u>Teorema 1</u>. El coeficiente $\bar{u}_0$ del estimador lineal no-sesgado de mínima varianza $\hat{\phi}_0 = \bar{u}_0^t \hat{b}$ de la funcional $\phi = \bar{c}^t \bar{x}$ está dado por

$$\bar{u}_0 = S^{-1} A(A^t S^{-1} A)^{-1} \bar{c}. \tag{16}$$

<u>Corolario 1</u>. Si $\bar{x}_0$ denota la solución del problema clásico de regresión lineal entonces la estimación no-sesgada y más eficiente de la k-ésima componente x_{k0} de $\bar{x}_0$ está dada por

$$\hat{x}_{k0} = \bar{e}_k^t (A^t S^{-1} A)^{-1} A^t S^{-1} \hat{b}. \tag{17}$$

<u>Corolario 2</u>. Si $\bar{x}_0$ denota la solución del problema clásico de regresión lineal entonces la estimación no-sesgada y más eficiente de $\bar{x}_0$ está dada por

$$\hat{x}_0 = (A^t S^{-1} A)^{-1} A^t S^{-1} \hat{b}. \tag{18}$$

(b). Una Aplicación a Sistemas de Ecuaciones Lineales Algebraicas.

Considérese el sistema lineal

$$(1) \qquad B\bar{x} = \bar{d}.$$

en donde B es una matriz real nxm con m>n y de rango n, $\bar{d}$ es un vector en R^n dado, y $\bar{x}$ un vector para encontrar en R^m.

En la siguiente discusión nosotros pensamos a B como una transformación lineal $B: R^m \to R^n$. En estos términos el sistema lineal dado en (1) siempre tiene solución para cualquier vector $\bar{d}$ en R^n porque $Im(B) = R^n$ debido a que el rango de B es n.

Ahora como m>n tenemos que el subespacio núcleo N(B) de B es distinto del subespacio $\{\bar{0}\}$ de R^m, de donde

$$\{\bar{x} \mid x = \bar{x}_p + \bar{n},\ B\bar{n}=0 \text{ y } B\bar{x}_p = \bar{d}\}$$

es el conjunto solución del sistema $B\bar{x}=\bar{d}$.

El problema que nos plantea la discusión del párrafo anterior consiste en dar un criterio que nos permita asociarle sin ambiguedades una única solución al sistema $B\bar{x}=\bar{d}$.

El problema de calcular un estimador no-sesgado y más eficiente de la solución del modelo clásico de regresión lineal nos sugiere el siguiente criterio para la solución de nuestro problema:

Asociarle como solución al sistema $B\overline{x}=\overline{d}$, la solución $\overline{x}_0$ de menor norma de $B\overline{x}=\overline{d}$.

Ahora como la norma de un vector $\overline{x}$ en R^m está dada por

$$||\overline{x}||^2 = \langle \overline{x},\overline{x} \rangle = \overline{x}^t\, \overline{x} \;,$$

se tiene por el teorema 1 de la subsección anterior para $S=I$ (la identidad en R^m) que la solución de menor norma $\overline{x}_0$ de $B\overline{x}=\overline{d}$ está dada por

$$(2) \qquad \overline{x}_0 = B^t(B\, B^t)^{-1}\, \overline{d}.$$

3. LA MEJOR SOLUCION APROXIMADA Y LA INVERSA GENERALIZADA DE UNA MATRIZ

Considérese el sistema lineal algebraico

$$(1) \qquad A\overline{x} = \overline{b}$$

en donde A es una matriz real mxn de rango r dada, $\overline{b}$ un vector en R^m dado y $\overline{x}$ un vector en R^n por encontrar.

Si (1) no tiene solución entonces siguiendo el criterio de los mínimos cuadrados, nosotros siempre le podemos asociar a (1) una solución aproximada $\bar{x}_a$. Esto es, $\bar{x}_a$ es una solución aproximada de (1) si

$$||A\bar{x}_a - \bar{b}||^2 \leq ||A\bar{x} - \bar{b}||^2$$

para toda $\bar{x}$ en R^n.

Obsérvese que si $\bar{x}_p$ es una solución de (1) entonces $\bar{x}_a = \bar{x}_p$.

Ahora si $r < \min\{n,m\}$, nosotros tenemos que (1) tiene una infinidad de soluciones, a saber cualquier $\bar{x}$ del conjunto

$$\{\bar{x} \mid \bar{x} = \bar{x}_a + \bar{n},\ A\bar{n} = \bar{0}\}$$

es también una solución aproximada de (1). Lo que nos lleva al problema de cómo seleccionar sin ambiguedades una solución aproximada de (1). El ejemplo de la sección anterior nos sugiere asociarle a (1) la solución aproximada de menor norma que la denotaremos por $\bar{x}_0$.

Obsérvese que si $r=n$ entonces $\bar{x}_a$ es única, y por lo tanto $\bar{x}_0 = \bar{x}_a$.

La discusión anterior nos sugiere introducir la siguiente

Definición 1. Diremos que $\bar{x}_0$ es la mejor solución aproximada (m.s.a.) de $A\bar{x}=\bar{b}$ si

(i) $||A\bar{x}_0-\bar{b}||^2 \leq ||A\bar{x}-\bar{b}||^2$,

para toda $\bar{x}$ en R^n. Y

(ii) Si $\bar{x}_a$ satisface $||A\bar{x}_a-\bar{b}||^2 \leq ||A\bar{x}-\bar{b}||$.

para toda $\bar{x}$ en R^n entonces $||\bar{x}_0|| \leq ||\bar{x}_a||$.

En la siguiente figura se ilustra a $\bar{x}_0$ la m.s.a. de (1) para el caso m=n=3 y r=2

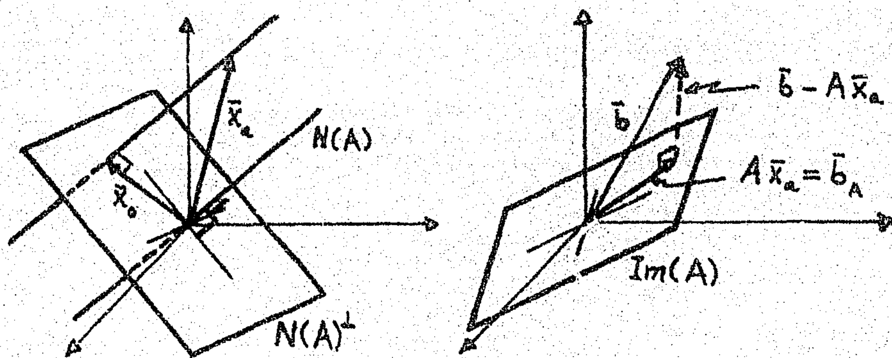

Figura 4. Ilustración de la m.s.a. $\bar{x}_0$ de $A\bar{x}=\bar{b}$ para el caso m=n=3 y r=2. N(A) denota el conjunto solución de $A\bar{x}=\bar{0}$.

En los casos que m=r y n=r, nosotros tenemos que

$$\bar{x}_0 = (A^tA)^{-1} A^t \bar{b}$$

Y

$$\bar{x}_0 = A^t(AA^t)^{-1} \bar{b}$$

respectivamente.

El párrafo anterior nos sugiere la existencia de una matriz A^+ de dimensiones nxm que llamaremos la <u>inversa generalizada</u> de A tal que

$$\bar{x}_0 = A^+ \bar{b}$$

nos da la m.s.a. de $A\bar{x}=\bar{b}$.

En el capítulo siguiente se prueba que A^+ existe, se calcula explícitamente y se da una caracterización de A^+.

BIBLIOGRAFIA

1. Ben Noble, "Applied Linear Algebra", Prentice Hall, 1969.

2. D.G. Luenberger, "Optimization by Vector Space Methods", Wiley and Sons, 1969.

3. B.W. Rust and W.R. Burnis, "Mathematical Programming and the Numerical Solution of Linear Equations", Elsevier, 1972.

CAPITULO II

LA DESCOMPOSICION SINGULAR
Y
LA INVERSA GENERALIZADA
DE UNA
MATRIZ

En este capítulo se establece una fórmula explícita de la matriz inversa generalizada A^+ con importantes implicaciones numéricas y una caracterización de la misma, via la descomposición singular de A.

1. LA DESCOMPOSICION ESPECTRAL

Considérese la ecuación lineal no-homogénea

$$(1) \qquad A\bar{x} = \bar{b},$$

en donde A es una matriz cuadrada de orden nxn y de elementos reales, $\bar{x}$ y $\bar{b}$ dos vectores en R^n.

Observemos que si $\bar{x} = P\,\bar{x}^1$ es una transformación de coordenadas de R^n dada por la matriz no-singular P entonces la resolución de (1) está completamente determinada por la resolución del sistema lineal dado por

$$(2) \qquad B\,\bar{x}^1 = \bar{b}^1$$

en donde

$$(3) \qquad B = P^{-1} A P,$$

y recíprocamente.

Ahora si A es una matriz real de la forma más sencilla, esto es, si A=D en donde

$$(4) \qquad D = \begin{bmatrix} d_1 & & 0 \\ & d_2 & \\ & & \ddots \\ 0 & & & d_n \end{bmatrix}$$

es una <u>matriz diagonal</u>. Entonces es inmediato ver que la resolución de (1) equivale a resolver las siguientes n ecuaciones escalares

$$(5) \qquad d_k x_k = b_k ,$$

para k=1,...,n que por ser independientes para su solución entre sí es fácil ver cuando (1) tiene solución, y si ésta es única o no.

Si para una transformación de coordenadas $\overline{x} = P \overline{x}^1$ de R^n en (3), B es una matriz diagonal entonces diremos que A es <u>diagonalizable</u>.

Por lo dicho hasta aquí, se sugiere investigar las condiciones bajo las cuales una matriz cuadrada A es diagonalizable.

Para ello observemos que D está caracterizada por tener la propiedad de que las n parejas $\{d_k,\bar{e}_k\}$ para $k=1,\ldots,n$ son todas las soluciones de la ecuación

$$D\bar{y} = \lambda\bar{y},$$

en donde todas las d_k son reales y los vectores $\bar{e}_k$ son los vectores que constituyen la "base canónica" de R^n.

Del párrafo anterior es natural preguntarse que si para una matriz cuadrada A, la ecuación

$$(6) \qquad A\bar{y} = \lambda\bar{y}$$

tiene n soluciones $\{d_k,\bar{p}_k\}$ con d_k en R para toda k y los vectores $\bar{p}_k$ para $k=1,\ldots,n$ (los cuales siempre podemos escoger unitarios) linealmente independientes (l.i.) entonces ¿será A diagonalizable?

Para responder a nuestra pregunta empecemos por considerar las hipótesis sobre A, esto es, que

$$(7) \qquad A\,\bar{p}_k = d_k\,\bar{p}_k\,,$$

para $k=1,\ldots,n$. Ahora si P es la matriz cuyas columnas son los vectores l.i. $\bar{p}_k$ entonces tenemos que las relaciones dadas en (7) implican que

$$(8) \qquad AP = PD$$

con P no-singular y D una matriz diagonal cuya diagonal principal está dada por los reales d_k. Luego entonces A es diagonalizable ya que de (8) se tiene que

$$(9) \qquad P^{-1} A P = D.$$

El recíproco de nuestra pregunta es también cierto, siendo inmediatamente verificable de (9) y (8).

De los dos párrafos anteriores tenemos que una matriz cuadrada A es diagonalizable si y sólo si existen n números reales $\lambda = d_k$ tales que el sistema lineal

$$(10) \qquad (A-\lambda I)\bar{y} = \bar{0}$$

tiene una solución unitaria $\bar{y} = \bar{p}_k$ para cada k, y en donde los vectores $\bar{p}_k$ son l.i. Pero esto último ocurre si y sólo si el polinomio "característico" de A

$$(11) \qquad p(\lambda) = \det(A-\lambda I)$$

con coeficientes reales y de grado n tiene n raices d_k (valores propios) reales y las correspondientes soluciones (vectores propios) unitarias $\bar{y}=\bar{p}_k$ de (10) son l.i.

Como es sabido, en general $p(\lambda)$ tiene n raices complejas, luego no toda matriz cuadrada A es diagonalizable.

Más aún, no basta con que todas las raices de p (λ) sean reales para que A sea diagonalizable, como se muestra en el

Ejemplo 1 La matriz

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

no es diagonalizable. En efecto, las raices del polinomio característico $p(\lambda) = \det(A-\lambda I) = (\lambda-1)^2$ son $\lambda_1 = \lambda_2 = 1$, y sus vectores propios correspondientes son paralelos al vector $(1,0)^t$. Luego A no es diagonalizable.

Proposición 1. Sea A una matriz cuadrada dada. Entonces si todos los valores propios de A son reales y distintos, A es diagonalizable.

Su demostración es bien conocida.

Hasta aquí, no hemos dado condiciones sobre una matriz A cuadrada bajo las cuales es diagonalizable. Para estudiar este punto, recordemos que nosotros en un principio estamos interesados en calcular la m.s.a. asociada a un sistema lineal que como ya vimos, tiene que ver ----- con la Geometría de R^n. Aqui, Geometría en R^n significa paralelismo, perpendicularidad y norma de vectores conceptos que, en R^n están dados en términos de su producto interior (usual)

$$\langle \bar{x}, \bar{y} \rangle = \bar{y}^t \, \bar{x}.$$

De aquí que, nosotros estamos interesados en investigar las condiciones sobre A bajo las cuales existe una transformación de coordenadas de R^n dada por una matriz cuadrada U que diagonalize a A y que conserve el producto interior de R^n.

A las matrices U que conservan el producto interior las llamaremos <u>ortogonales</u>.

<u>Proposición 2</u>. Sea U una matriz cuadrada. Entonces las siguientes expresiones son equivalentes

(i) U es ortogonal.

(ii) $U \, U^t = U^t \, U = I$. Esto es, $U^{-1} = U^t$.

(iii) Los vectores columna $\bar{u}_k$ de U constituyen una base β_u "ortonormal" de V_n. Esto es,

$$\langle \bar{u}_i, \bar{u}_j \rangle = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j. \end{cases}$$

Su demostración es totalmente directa.

Si A es diagonalizable por una transformación de coordenadas ortogonal U de R^n entonces diremos que A es <u>diagonalizable ortogonalmente</u>.

Ahora supongamos que A es diagonalizable ortogonalmente. Esto es, existe una matriz U ortogonal tal que $U^t AU = D$, en donde D es una matriz diagonal. Y como $D^t = D$ tenemos que $U^t A^t U = U^t A U$, y por lo tanto $A^t = A$.

Hemos probado así la siguiente

<u>Proposición 3</u>. Una condición necesaria para que una matriz cuadrada A sea diagonalizable ortogonalmente es que A sea "simétrica", esto es que $A^t = A$.

De este último resultado cabe sospechar que la simetría de A es también una condición suficiente para que A sea diagonalizable ortogonalmente como en efecto lo muestran los siguientes teoremas bien conocidos:

<u>Teorema 1</u>. Si A es simétrica entonces todos los valores propios de A son reales. Más aún, si d_i y d_j son dos valores propios de A distintos para $i \neq j$ entonces sus vectores propios correspondientes $\bar{u}_i$ y $\bar{u}_j$ son mutuamente ortogonales.

<u>Teorema 2</u>. Si A es simétrica entonces existen n vectores

propios y unitarios $\bar{u}_k$ ortogonales entre sí correspondientes a los n valores propios $\lambda = d_k$ no necesariamente distintos.

Para la demostración de estos dos teoremas ver [1].

Por último de la Proposición 3 y el Teorema 2 se sigue el siguiente

Teorema 3. (de la Descomposición Espectral). Sea A una matriz cuadrada y real. Entonces A es simétrica si y sólo si existe una matriz ortogonal U tal que A admite la descomposición

$$(12) \qquad A = U\,D\,U^t$$

en donde D es una matriz diagonal.

2. LA DESCOMPOSICION SINGULAR.

Supongamos que se nos da el sistema lineal

$$(1) \qquad A\,\bar{x} = \bar{b},$$

en donde A es una matriz cuadrada que no es simétrica (i.e. $A \neq A^t$). Luego, por la sección anterior sabemos que (1) no se puede llevar a un sistema lineal diagonal

$$(2) \qquad D\,\bar{x}^1 = \bar{b}^1,$$

mediante una transformación ortogonal $\overline{y}=U\,\overline{y}^{1}$ de coordenadas de R^n. Dicho de otro modo, A no admite la descomposición

$$(3) \qquad A = U\,D\,U^t$$

para alguna matriz U ortogonal.

Sin embargo, ¿admitirá A una descomposición similar a la dada por (3)? Para contestar esta pregunta, tomemos en cuenta que una matriz B admite la descomposición espectral siempre y cuando $B=B^t$. Y como nosotros tenemos que $A\neq A^t$, lo anterior nos sugiere considerar el sistema lineal aumentado

$$(4.a) \qquad A\,\overline{x} = \overline{b}$$

$$(4.b) \qquad A^t\,\overline{y} = \overline{c}$$

en vez de considerar aisladamente al sistema lineal dado por (1), observando que las soluciones de (4.a) y (4.b) son independientes una de la otra.

Pero a (4.a) y (4.b) se les puede escribir como un solo sistema

$$(5) \qquad S\,\overline{z} = \overline{f}$$

en donde

$$(6) \qquad S = \begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix}, \quad \overline{z} = \begin{bmatrix} \overline{y} \\ \overline{x} \end{bmatrix}, \quad y \ \overline{f} = \begin{bmatrix} \overline{b} \\ \overline{c} \end{bmatrix}$$

Ahora como $S=S^t$, los resultados de la sección anterior son aplicables a (5). Esto es, existen $2n$ parejas $\{d_k, \bar{w}_k\}$ con d_k en los reales, el conjunto $\{\bar{w}_k\}$ ortonormal en R^{2n} y tales que

$$(7) \qquad S\bar{w}_k = d_k \bar{w}_k .$$

Si $\bar{w}_k = \begin{bmatrix} \bar{v}_k \\ \bar{u}_k \end{bmatrix}$ con $\bar{u}_k$ y $\bar{v}_k$ en R^n entonces de (6) tenemos que (7) equivalente a las relaciones

$$(8.a) \qquad A\,\bar{u}_k = d_k\,\bar{v}_k$$

y

$$(8.b) \qquad A^t\,\bar{v}_k = d_k\,\bar{u}_k ,$$

para $k=1,2,\ldots,2n$.

Ahora si U y V son las matrices formadas por los conjuntos de vectores $\{\bar{u}_k\}$ y $\{\bar{v}_k\}$ respectivamente, nosotros tenemos de (8.a) y (8.b) que

$$(9.a) \qquad AU = VD$$

y

$$(9.b) \qquad A^t\,V = UD,$$

en donde D es la matriz diagonal de orden 2n x 2n cuya diagonal principal está dada por las d_k.

Para eliminar a U y a V de (9.a) y (9.b) y así obtener unas descomposiciones similares a la espectral para A y A^t respectivamente se requiere que ambas matrices U y V sean en primer término cuadradas y en segundo ortogonales.

O sea, nuestro problema es ver que siempre es posible seleccionar n tercias $\{d_k, \bar{u}_k, \bar{v}_k\}$ que satisfagan (8.a) y (8.b) y tales que los conjuntos $\{\bar{u}_k\}$ y $\{\bar{v}_k\}$ sean ambos ortogonales.

Para ello eliminemos a $\bar{v}_k$ y $\bar{u}_k$ de (8.a) y (8.b) multiplicando a (8.a) y (8.b) por A^t y A respectivamente, obteniendo que

$$A^t A \bar{u}_k = d_k^2 \bar{u}_k \tag{11.a}$$

y

$$A A^t \bar{v}_k = d_k^2 \bar{v}_k \, . \tag{11.b}$$

Ahora ambas matrices $A^t A$ y AA^t son simétricas, no-negativas definidas y con los mismos valores propios. Las dos primeras afirmaciones son evidentes. Para demostrar la tercera, sea λ un valor propio de A^tA, esto es, $A^tA\,\bar{u} = \lambda\bar{u}$, de donde $AA^t(A\bar{u}) = \lambda(A\bar{u})$, lo que muestra que λ es un valor propio de AA^t; de manera forma $\{0,\bar{0},\bar{v}_k\}$ para $k=r+1,\ldots,n$.

Esto es, si U y V son las matrices cuyas k-ésimas columnas son u_k y v_k para $k=1,\ldots,n$, respectivamente. Entonces U y V son ortogonales y satisfacen

$$(13.a) \qquad A\,U = V\,\Delta$$

y

$$(13.b) \qquad A^t V = U\,\Delta,$$

o bien

$$(14.a) \qquad A = V\,\Delta\,U^t$$

y

$$(14.b) \qquad A^t = U\,\Delta\,V^t\,,$$

en donde la matriz Δ de orden nxn es de la forma

$$(15) \qquad \Delta = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

con D_r una sub-matriz diagonal de orden rxr, en la cual los elementos de su diagonal principal d_k se pueden tomar siempre positivos.

Por último, observemos que los argumentos de toda la discusión anterior se sostienenpara el caso de una matriz rectangular A de orden mxn bajo cambios apropiados de dimensión.

Teniendo en mente el párrafo anterior, la discusión de esta sección se puede resumir en el siguiente

Teorema 1. (de la Descomposición Singular) Si A una matriz de dimensiones mxn con elementos reales entonces siempre existen dos matrices ortogonales U y V de órdenes nxn y mxm respectivamente y tales que A y A^t admiten las descomposiciones singulares

$$A = V \Delta U^t$$

y

$$A^t = U \Delta^t V^t,$$

en donde las matrices Δ y Δ^t de órdenes mxn y nxm son de la forma

$$\Delta = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

y

$$\Delta^t = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

respectivamente, con D_r una submatriz diagonal de orden r en la cual los elementos de su diagonal principal se pueden tomar siempre positivos.

3. LA INVERSA GENERALIZADA DE UNA MATRIZ.

Es bien sabido que si A es una matriz nxn y de rango n, entonces la solución del sistema lineal

$$(1) \qquad A\bar{x} = \bar{b}$$

existe, es única y está dada por

$$(2) \qquad \bar{x} = A^{-1}\,\bar{b}$$

para cualquier $\bar{b}$ en R^n dada. Aqui, A^{-1} denota la matriz inversa de A.

En cambio, si A es de rango r con $1 \leq r < n$ entonces (1) no siempre tiene solución y cuando tiene, tiene una infinidad de soluciones.

Ahora si A es una matriz rectangular mxn y de rango r con $1 \leq r \leq \min\{n,m\}$ entonces (1) no siempre tiene solución y cuando tiene, tiene una infinidad. Como puede observarse fácilmente el caso del párrafo anterior queda incluído en el presente. Luego para no ser redundantes nos limitaremos a discutir el caso general.

De aquí y en lo que resta de esta sección, A es una matriz rectangular mxn de rango r menor que el min $\{n,m\}$

y tal que (1) no tiene solución.

Aunque (1) no tenga una solución tiene sin embargo sentido calcular como ya se vió,

La Mejor Solución Aproximada asociada a (1): nosotros diremos que $\overline{x}_0$ en R^n es la m.s.a. de (1) si $\overline{x}_0$ satisface

(3.a) $||A\overline{x}_0-\overline{b}||^2 \leq ||A\overline{x}-\overline{b}||^2$, para toda $\overline{x}$ en R^n

y

(3.b) Si $\overline{x}_1$ es tal que $||A\overline{x}_1-\overline{b}||^2 \leq ||A\overline{x}-\overline{b}||^2$ para toda $\overline{x}$ en R^n entonces $||\overline{x}_1||^2 \geq ||\overline{x}_0||^2$.

Es decir $\overline{x}_0$ es la $\overline{x}$ de menor norma que minimiza a $||A\overline{x}-\overline{b}||^2$. No es difícil ver que $\overline{x}_0$ siempre existe y que además es única como una consecuencia del

Teorema 1. (de la Proyección. Caso finito). Sea H un subespacio propio de R^n y sea $\overline{b}$ en R^n-H. Entonces existe un único vector $\overline{y}_0$ en H tal que $||\overline{y}_0-\overline{b}||^2 \leq ||\overline{y}-\overline{b}||^2$ para toda $\overline{y}$ en R^n. Más aún, $\overline{y}_0$ minimiza a $||\overline{y}-\overline{b}||^2$ si y sólo si $\langle \overline{b}-\overline{y}_0, \overline{y} \rangle = 0$ para toda $\overline{y}$ en H.

Su demostración es directa.

Ahora ¿será posible asociarle a A una matriz B tal que $\overline{x}_0=B\overline{b}$ sea la m.s.a. de (1)?

Para ello, vemos primero que por el teorema de la descomposición singular tenemos que

$$\begin{aligned} ||Ax-b||^2 &= ||V(\Delta U^t\overline{x}-V^t\overline{b})||^2 \\ &= ||\Delta U^t\,\overline{x}-V^t\overline{b}||^2 \\ &= ||\Delta\overline{z}\;-\overline{c}||^2, \end{aligned}$$

en donde $\overline{z} = U^t\,\overline{x}$ y $\overline{c} = V^t\,\overline{b}$.

Y como U,U^t,V,V^t conservan la norma tenemos que encontrar la m.s.a. de (1) es equivalente a buscar la m.s.a. de

$$(4)\qquad \Delta\overline{z} = \overline{c}.$$

Para calcular la m.s.a. de (4) calculemos primero las $\overline{z}$ que minimizan a $||\Delta\overline{z}-\overline{c}||^2$. Aquí,

$$(5)\quad ||\Delta z-c||^2 = \sum_{k-1}^{r} (d_k z_k-c_k)^2 + \sum_{k=r+1}^{n} c_k{}^2$$

De aquí que, las $\overline{z}$ que minimizan a (5) son aquéllas que satisfacen las r ecuaciones escalares

$$d_k z_k=c_k,\quad k=1,\ldots,r.$$

De donde las $\bar{z}$ que minimizan a (5) están dadas por

$$(6) \qquad (\bar{z})^t = (\frac{c_1}{d_1}, \ldots, \frac{c_r}{d_r}, \alpha_{r+1}, \ldots, \alpha_n)$$

para α_k, $k=r+1,\ldots,n$ escalares arbitrarios.

Y por lo tanto la m.s.a. $\bar{z}_0$ de (4) está dada por

$$(7) \qquad \bar{z}_0 = \begin{bmatrix} c_1/d_1 \\ c_2/d_2 \\ \vdots \\ c_r/d_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = D^+ \bar{c}$$

en donde D^+ es la matriz de dimensiones nxm

$$(8) \qquad D^+ = \left[\begin{array}{cccc|c} d_1^{-1} & & & & \\ & d_2^{-1} & & & 0 \\ & & \ddots & d_r^{-1} & \\ \hline & & 0 & & 0 \end{array}\right]$$

Ahora llevando (7) a nuestro problema inicial, esto es, si $\bar{z}_0=U^t\,\bar{x}_0$ y $\bar{c}=V^t\,\bar{b}$ entonces tenemos que

$$U^t\,\bar{x}_0 = D^+\,V^t\,\bar{b}$$

o bien que

$$(9) \qquad \bar{x}_0 = A^+\,\bar{b}\ ,$$

en donde

(10) $$A^{+} = U\,\Delta^{+}\,V^{t}$$

la cual llamaremos la matriz <u>inversa generalizada</u> asociada a A, la que claramente es única.

De (9) se ve fácilmente que A^{+} viene jugando un papel muy análogo al de A^{-1}. Más aún $A^{+} = A^{-1}$ cuando A^{-1} existe.

Del hecho que $A = V\Delta U^{t}$ es directo ver que los subespacios núcleo $N(A)$ e imagen $Im(A)$ de A están dados por los vectores $\bar{u}_{r+1},\ldots,\bar{u}_n$ y $\bar{v}_1,\ldots,\bar{v}_r$ respectivamente. Y de (10) se tiene que los subespacios núcleo $N(A^{+})$ e imagen $Im(A^{+})$ de A^{+} están dados por $\bar{v}_{r+1},\ldots,\bar{v}_m$ y $\bar{u}_1,\ldots,\bar{u}_r$. De aquí se sigue que $N(A)\perp Im(A^{+})$ y que $Im(A)\perp N(A^{+})$.

En base al párrafo anterior tenemos la siguiente caracterización de la inversa generalizada A^{+} de A.

<u>Teorema 2</u>. A^{+} es la única matriz que satisface las siguientes dos ecuaciones

$$A\,X = P_{Im(A)} \quad y \quad X\,A = P_{N(A)^{\perp}} \;.$$

BIBLIOTECA CENTRAL
U.N.A.M.

en donde $P_{Im(A)}$ y $P_{N(A)^{\perp}}$ denotan las proyecciones ortogonales sobre los subespacios $Im(A)$ y $N(A)^{\perp}$ respectivamente.

De la descomposición singular de A y de (10) tenemos que

$$AA^{+} = V\Delta U^{t}U\Delta^{+}V^{t} = V\Delta\Delta^{+}V^{t}$$

$$= V \begin{bmatrix} I_r & 0 \\ & \\ 0 & 0 \end{bmatrix} V^{t} = \begin{bmatrix} I_r & 0 \\ & \\ 0 & 0 \end{bmatrix}$$

en donde I_r denota la matriz identidad de orden r. Y de aquí, y por el párrafo inmediatamente anterior del enunciado del teorema es inmediato ver que

$$A\,A^{+} = P_{Im(A)}\,.$$

De manera completamente análoga se tiene que

$$A^{+}\,A = P_{N(A)^{\perp}}.$$

Nótese que AA^{+} y $A^{+}A$ se parecen mucho a las matrices identidad sobre R^{m} y R^{n} respectivamente. Lo que justifica el denominativo de A^{+}.

Ahora supongamos que B es otra matriz que satisface

$$BA = P_{N(A)^{\perp}} \quad y \; AB=P_{Im(A)}\,.$$

De la segunda igualdad de arriba se tiene que

$$(AB - AA^{+})\bar{b} = \bar{0},$$

para cualquier $\bar{b}$ en R^m. Pero como $A(B-A^{+}) = AB-AA^{+}$ tenemos que $(B-a^{+})\bar{b}$ está en el $N(A)$. Esto es,

$$B\bar{b} = A^{+}\bar{b}+\bar{n}$$

con $\bar{n}$ en el $N(A)$. Nosotros afirmamos que $\bar{n}=\bar{0}$, ya que de lo contrario $B\bar{b}$ no estaría en $N(A)^{\perp}$ contradiciendo $Im(B) = Im(BA) = Im(P_{N(A)^{\perp}}) = N(A)^{\perp}$.

4. "ESTABILIDAD" EN LA MEJOR SOLUCION APROXIMADA DE LOS SISTEMAS LINEALES $A\bar{x}=\bar{b}$ y $A^tA\bar{x}=A^t\bar{b}$

Considérese el sistema lineal

$$(1) \qquad A\bar{x} = \bar{b}$$

en donde A es una matriz real de dimensiones mxn conocida y $\bar{b}$ un vector en R^m dado. Se dice en lenguaje llano que (1) es un sistema lineal estable si a "pequeñas" variaciones de $\bar{b}$ se tienen "pequeñas" variaciones en su m.s.a. $\bar{x}_0$.

Nuestro propósito en esta sección es calcular una "medida" así como su comparación de la estabilidad de los sistemas lineales $A\bar{x}=\bar{b}$ y $A^tA\bar{x} = A^t\bar{b}$.

Para no distraer la atención de nuestro objetivo, nos permitiremos el uso sin previa discusión de conceptos que caen en el campo de estudio del Análisis Numérico.

Una discusión detallada al respecto se encuentra en [1].

Regresemos a nuestro problema en lenguaje técnico.
Si $\bar{x}_1$ es la m.s.a. del sistema lineal perturbado

$$(2) \qquad A\,\bar{x} = \bar{b} + \bar{\varepsilon}$$

entonces el vector variación en la solución de (1) viene dada por

$$\bar{x}_\varepsilon = \bar{x}_1 - \bar{x}_0 = A^+(\bar{b}+\bar{\varepsilon}) - A^+\bar{b} = A^+\bar{\varepsilon}$$

de donde "el coeficiente condicional relativo" $K(A,\bar{b})$ de (1) es la menor constante $k>0$ tal que

$$(3) \qquad \frac{||\bar{x}_\varepsilon||}{||\bar{x}_0||} \qquad \frac{||A^+\bar{\varepsilon}||}{||A^+\bar{b}||} \leq K\,\frac{||\bar{\varepsilon}||}{||\bar{b}||}.$$

Para calcular $K(A,\bar{b})$ veamos que (3) nos sugiere calcular la menor constante $M>0$ tal que

$$(4) \qquad ||A^+\bar{\varepsilon}|| \leq M\,||\bar{\varepsilon}||$$

para toda $\bar{\varepsilon}$ en R^m. Y la mayor constante $m>0$ tal que

(5) $$||A^{+}\,\overline{b}|| \geq m||\overline{b}||$$

para toda $\overline{b}$ en $R^{m} - N(A^{t})$.

Para calcular a M observemos primero que

$$||A^{+}\,\overline{\varepsilon}|| = ||U\Delta^{+}V^{t}\overline{\varepsilon}|| = ||\Delta^{+}V^{t}\overline{\varepsilon}||,$$

luego si $\overline{\eta} = \overline{V}^{t}\overline{\varepsilon}$ tenemos que $||\overline{\eta}|| = ||\overline{\varepsilon}||$. Esto es, la menor $M>0$ que satisface a (4) es la menor $M>0$ que satisface

(6) $$||\Delta^{+}\,\overline{\eta}|| \leq M||\overline{\eta}||.$$

Como siempre podemos suponer que en la submatriz diagoanl D_r de Δ^{+}, $d_r^{-1} \geq d_{r-1}^{-1} \geq \ldots \geq d_1^{-1} > 0$, tenemos que

$$||\Delta^{+}\overline{\eta}||^2 = \sum_{k=1}^{r} (d_k^{-1}\eta_k)^2 \leq \sum_{k=1}^{r} (d_r^{-1})^2\ \eta^2_k$$

$$\leq (d_r^{-1})^2 \sum_{k=1}^{m} \eta_k^2 = (d_r^{-1})^2\ ||\overline{\eta}||^2.$$

Y por lo tanto

(7) $$||A^{+}\overline{\varepsilon}|| \leq d_r^{-1}\,||\overline{\varepsilon}||,$$

para toda $\overline{\varepsilon}$ en R^{m}. O bien que $||A^{+}\,\overline{y}|| \leq d_r^{-1}$ para toda $\overline{y}$ en R^{m} con $||\overline{y}|| = 1$. Más aún,

$$\sup_{\|\bar{y}\|=1} \|A^{+}\bar{y}\| = d_r^{-1}$$

ya que $\|A^{+}\bar{v}_r\| = d_r^{-1}$. Los tres párrafos últimos anteriores son la demostración esencialmente de la siguiente

<u>Proposición 1</u>. La "norma uniforme" $\|A\| = \sup_{\|\bar{x}\|=1} \|A\bar{x}\|$ con respecto a las normas "euclidianas" de R^n y R^m respectivamente de la matriz real A de dimensiones mxn es igual al mayor en valor absoluto de sus valores singulares.

Consecuentemente la menor constante $M>0$ que satisface a (4) es d_r^{-1}.

Para calcular la mayor constante $m>0$ cumpliendo con (5) observemos que

$$(8) \qquad \|\bar{b}_A\| = \|P_{Im(A)}\,\bar{b}\| = \|A\,A^{+}\,\bar{b}\|\cdot$$

De (8) y la Proposición 1 se sigue que

$$(9) \qquad \|\bar{b}_A\| \le d_1 \|A^{+}\,\bar{b}\| \;,$$

o bien que

$$(10) \qquad \|A^{+}\bar{b}\| \ge \frac{\|\bar{b}_A\|}{d_1} = d_1^{-1}\,\frac{\|\bar{b}_A\|}{\|\bar{b}\|}\;\|\bar{b}\|.$$

Y por lo tanto, la mayor constante m>0 que satisface a a (5) es $d_1^{-1}\dfrac{||\overline{b}_A||}{||\overline{b}||}$ ya que d_1 es la mayor constante que satisface a (9).

Ahora de (7) y (10) se tiene que

$$\frac{||\overline{x}_\varepsilon||}{||\overline{x}_0||} \leq (d_1 d_r^{-1} \frac{||\overline{b}||}{||\overline{b}_A||}) \frac{||\overline{\varepsilon}||}{||\overline{b}||}$$

Esto es, el coeficiente condicional relativo de (1) viene dado por

$$K(A,\overline{b}) = ||A||\ ||A^+||\ \frac{||\overline{b}||}{||\overline{b}_A||}$$

La discusión anterior se resume en el siguiente

<u>Teorema 1</u>. El coeficiente condicional relativo $K(A,\overline{b})$ del sistema lineal $A\overline{x}=\overline{b}$ está dado por

$$K(A,\overline{b}) = K(A)\cdot K(\overline{b}) \qquad (11)$$

donde $K(A) = ||A||\ ||A^+||$ y $K(\overline{b}) = ||\overline{b}||\ /\ ||\overline{b}_A||$.

Como $P_{Im(A)} = A\,A^+$ se tiene que $K(A) \geq 1$. Por el teorema de Pitágoras también se tiene que $K(\overline{b}) \geq 1$. Y por lo tanto, $K(A,\overline{b}) \geq 1$.

En la práctica generalmente se resuelve la ecuación normal

$$(12) \qquad A^t A \bar{x} = A^t \bar{b}$$

asociada al sistema lineal $A\bar{x}=\bar{b}$ en vez de este último por las razones siguientes: primera, (12) es siempre soluble porque $\mathrm{Im}(A^tA) = \mathrm{Im}(A^t)$; segunda, la solución de menor norma de (12) es la m.s.a. de (1); y tercera, A^tA es una matriz cuadrada y simétrica.

El inconveniente de resolver el sistema $A\bar{x}=\bar{b}$ via su ecuación normal $\qquad A^tA\bar{x} = A^t\bar{b}$
es que frecuentemente en las aplicaciones este último es menos estable. Dicho de otro modo, en las aplicaciones generalmente $K(A,\bar{b}) \leq K(A^tA,\bar{b})$.

Antes de hacer ver lo dicho en el párrafo anterior, calculemos $K(A^tA,\bar{b})$. Procediendo análogamente como se hizo para calcular $K(A,\bar{b})$, se tiene que

$$\frac{||(A^tA)^+ A^t \bar{\varepsilon}||}{||\bar{x}_0||} \left(= \frac{||(A^tA)^+ A^t \bar{\varepsilon}||}{||(A^tA)^+ A^t \bar{b}||}\right) \leq \frac{||(A^tA)^+|| \; ||A^t|| \; ||\bar{\varepsilon}||}{||(A^tA)||^{-1}||A^t \bar{b}||}$$

de donde

$$(13) \qquad K(A^tA,\bar{b}) = K(A)^2 ||\bar{A}|| \frac{||\bar{b}||}{||A^t \bar{b}||} .$$

En las aplicaciones generalmente el vector b está "cercano" al subespacio Im(A) como puede verse en el capítulo I, de donde se sigue que

$$(14) \qquad K(A,\bar{b}) = K(A)K(\bar{b}) \approx K(A)$$

ya que $K(\bar{b}) = ||\bar{b}|| \,/\, ||\bar{b}_A|| \approx 1$.

De (13) es inmediato obtener que

$$(15) \qquad K(A^tA,\ \bar{b}) \geq K(A)^2$$

para cualquier vector $\bar{b}$ en $R^m - K(A^t)$. Lo que muestra que $K(A^tA,\ \bar{b}) \geq K(A,\ \bar{b})$ ya que $K(A) \geq 1$.

BIBLIOGRAFIA

1. Ben Noble, "Applied Linear Algebra", Prentice Hall, 1969.

2. C. Lanczos, "Linear Differential Operators", Van Nostrand, 1964.

CAPITULO III

LA INVERSA GENERALIZADA PARA TRANSFORMACIONES LINEALES Y CONTINUAS ENTRE ESPACIOS DE HILBERT

1. INTRODUCCION

Es muy frecuente en Matemáticas introducir nuevos conceptos así como los procedimientos y técnicas para su estudio por analogía a conceptos, procedimientos y técnicas con los cuales ya estamos familiarizados.

En este capítulo vamos a extender el estudio de la Inversa Generalizada a transformaciones lineales y continuas $A: H_1 \to H_2$ entre dos espacios de Hilbert H_1 y H_2, las cuales son muy parecidas a las transformaciones lineales entre espacios vectoriales de dimensión finita.

La generalización descrita en el párrafo anterior es sugerida entre otras cosas por las ecuaciones integrales de Fredholm inconsistentes

$$(1) \qquad \int_a^b f(t,s)\, x(s)ds = b(t) \quad (a \le t \le b)$$

y

$$(2) \qquad \lambda x(t) + \int_a^b f(t,s)\, x(s)ds = b(t) \quad (a \le t \le b)$$

de primera y segunda clase respectivamente, en donde $f(t,s)$ y $b(t)$ son funciones dadas y $x(s)$ es una función por encontrar, todas ellas con valores reales.

2. ESPACIOS DE HILBERT

Consideremos la ecuación integral de Fredholm de primera clase

$$(1) \qquad \int_a^b f(t,s)x(s)ds = b(t) \quad (a \le t \le b)$$

en donde, por razones de existencia de la integral, a $f(t,s)$, $x(s)$ y $b(t)$ las supondremos por el momento continuas en sus respectivos dominios de definición.

Si (1) no tiene solución es natural para nosotros intentar calcular una "solución aproximada" de (1) según el criterio de los mínimos cuadrados. Esto es, calcular una función $x_a(\cdot)$ tal que minimice a la función

$$(2) \qquad \phi(x(\cdot)) = \left|\left| \int_a^b f(\cdot,s)x(s)ds - b(\cdot) \right|\right|^2$$

definida sobre el espacio vectorial de las funciones reales y continuas en $[a,b]$, que de aquí en lo sucesivo denotaremos por $C[a,b]$.

Pero como nosotros queremos conservar la imagen geométrica del criterio de los mínimos cuadrados expuesto en el Capítulo I, necesitamos que la norma usada en la definición de ϕ esté dada en términos de un producto interior para $C[a,b]$. Es clásico equipar a $C[a,b]$ con el producto interior

$$(3) \qquad \langle x(\cdot),y(\cdot)\rangle = \int_a^b x(s)y(s)ds$$

sugerido por el producto interior natural en R^n

$$\langle \bar{x},\bar{y}\rangle = \bar{y}^t\ \bar{x} = \sum_{k=1}^{n} x_k y_k \ .$$

Antes de pasar al problema de la existencia de una función $x_a(\cdot)$ en $C[a,b]$ que minimice a (2) con la norma definida por (3), se requiere que el conjunto sobre el cual se va a minimizar a ϕ, esto es, $C[a,b]$, sea completo con la norma definida por (3).

Un espacio vectorial normado V se dice que es <u>completo</u> si para toda sucesión $\{x_n\}$ en V con $||x_n-x_m||\to 0$ cuando $n,m\to\infty$ implica la existencia de una x en V tal que $||x_n-x||\to 0$ cuando $n\to\infty$.

No es difícil dar un ejemplo como puede verse en [2] con el cual se muestre que $C[a,b]$ no es un espacio vectorial completo con respecto a la norma definida por (3). Lo que nos lleva a completar a $C[a,b]$ con la norma definida por (3), se puede probar como puede verse en [2] que el espacio deseado es el espacio vectorial de las funciones reales cuadrado integrables según Lebesgue, que denotaremos por $L^2[a,b]$.

$L^2[a,b]$ es un ejemplo típico de una clase de espacios vectoriales normados que introducimos en la siguiente

<u>Definición 1</u>. Sea H un espacio vectorial real con producto interior, se dice que H es un espacio de <u>Hilbert</u> si éste es completo con respecto a la norma $||\cdot||^2 = \langle\cdot,\cdot\rangle$. Un espacio de Hilbert es <u>separable</u> si contiene un subconjunto denso numerable.

De aquí en lo sucesivo H denotará un espacio de Hilbert separable

Como ejemplos de espacios de Hilbert separables citamos los siguientes: 1) R^n con su producto interior usual; 2) L^2 el espacio vectorial de las sucesiones reales $\{x_k\}$ tales que $\sum_{k=1}^{\infty} x_k^2 < \infty$ con el producto interior $x,y = \sum_{k=1}^{\infty} x_k y_k$ si $x = \{x_k\}$ y $y = \{y_k\}$; los espacios $L^2(I)$ de las funciones reales cuadrado integrables según

Lebesgue definidas sobre un intervalo real I acotado o no, cerrado o no.

No es difícil probar que los siguientes resultados válidos para R^n siguen siendo válidos para espacios de Hilbert.

Proposición 1. (Desigualdad de Cauchy-Schwarz) Para todo x,y en H, $| x,y | \leq ||x|| \, ||y||$. La igualdad se satisface si y sólo si $x=\lambda y$ o $y=\theta$.

Corolario 1. (Desigualdad del Triángulo) Para todo x,y en H, $||x+y|| \leq ||x|| + ||y||$

Proposición 2. (Ley del Paralelogramo). Para todo x,y en H,

$$||x+y||^2 + ||x-y||^2 = (||x||^2 + ||y||^2).$$

Proposición 3. (Continuidad del Producto Interior) Si $x_n \to x$ y $y_n \longrightarrow y$ en H, entonces $\langle x_n, y_n \rangle \longrightarrow \langle x,y \rangle$. Consecuentemente, si $x_n \longrightarrow x$ entonces $||x_n|| \longrightarrow ||x||$.

La idea geométrica de ortogonalidad en R^n nos permite introducir la siguiente definición, por similaridad, en H, la cual juega un papel central en el presente desarrollo.

<u>Definición 2.</u> Dos vectores x e y en H se dice que son ortogonales si y sólo si $\langle x,y\rangle=0$, hecho que denotaremos por $x\perp y$. De manera análoga, diremos que x es ortogonal a un subconjunto S de H ($x\perp S$) si $x\perp y$ para toda y en S.

<u>Proposición 4.</u> (Teorema de Pitagoras) Si $x\perp y$ en H entonces $||x+y||^2 = ||x||^2 + ||y||^2$.

Su demostración es directa.

En el capítulo anterior vimos que la existencia así como la unicidad de una $\bar{x}_a$ en R^n que minimice a $||A\bar{x}-\bar{b}||$ está dada por el Teorema de la Proyección en R^n. Es natural preguntar si este resultado es válido para espacios de Hilbert. La respuesta es afirmativa y la da el siguiente

<u>Teorema 1</u> (de la Proyección). Sea H un espacio de Hilbert, M un subespacio propio de H y b un vector en H-M entonces un x_0 en M satisface $||x_0-b|| = \min\limits_{x \text{ en } M} ||x-b||$ cuando y sólo cuando $b-x_0\perp M$. Si además M es cerrado en H entonces para toda b en H existe un único vector x_0 en M tal que

$$||x_0-b|| = \min_{x \text{ en } M} ||x-b||.$$

Su demostración es clásica y puede verse en [3].

3. TRANSFORMACIONES LINEALES CONTINUAS, COMPACTAS Y CERRADAS.

(a). Transformaciones Lineales Continuas.

En la sección anterior se dieron varias razones por las cuales conviene considerar a la ecuación integral

$$(1) \qquad \int_a^b f(t,s)x(s)ds = b(t) \qquad (a \leq t \leq b)$$

con $x(\cdot)$ y $b(\cdot)$ en el espacio de Hilbert $L^2[a,b]$.

En el caso de sistemas lineales algebráicos $A\overline{x}=\overline{b}$ tenemos que la matriz A define siempre una transformación lineal y continua. La continuidad de A es una consecuencia del hecho que siempre existe una constante $K>0$ tal que $||A\overline{x}|| \leq K||\overline{x}||$ para toda $\overline{x}$, así como de la linealidad de A.

Nuestro problema ahora es ver bajo qué condiciones la ecuación integral (1) nos define una transformación lineal y continua $K\colon L^2[a,b] \to L^2[a,b]$. La manera natural de definir a K es por

$$(2) \qquad K(x(\cdot)) = \int_a^b f(\cdot,s)x(s)ds,$$

la cual claramente es lineal.

En el siguiente ejemplo se muestra que no toda transformación lineal $A:H_1 \rightarrow H_2$ es continua.

Ejemplo 1. Considérese la transformación

$$A:L^2[0,\infty) \rightarrow L^2[0,\infty)$$

definida por $A(x(\cdot)) = y(\cdot)$ en donde

$$y(t) = t\, x(t) \qquad (0 \leq t < \infty)$$

Es inmediato verificar que A es lineal.

Consideremos la sucesión $\{x_n(\cdot)\}$ en $L^2[0,\infty)$ definida por

$$x_n(t) = \begin{cases} \sqrt{3}/n & 0 \leq t \leq n \\ 0 & t > n \end{cases}$$

Ahora $x_n(\cdot) \rightarrow 0(\cdot)$ ya que

$$||x_n(\cdot)-0||^2 = \int_0^\infty |x_n(t)|^2 dt = \int_0^n \left(\frac{\sqrt{3}}{n}\right)^2 dt = \frac{3}{n}$$

Por otro lado,

$$||Ax_n(\cdot)|| = \int_0^\infty |tx_n(t)|^2 dt = \int_0^n \frac{3}{n^2}\, t^2 dt = n \geq 1$$

para toda $n \geq 1$. Y por lo tanto, A no es continua.

Sin embargo es directo demostrar, usando la desigualdad de Cauchy-Schwarz

$$\left(\int_a^b x(t)y(t)dt\right)^2 \leq \left(\int_a^b |x(t)|^2 dt\right) \left(\int_a^b |y(t)|^2 dt\right)$$

para $L^2[a,b]$, que si

$$N^2 = \int_a^b \int_a^b |f(t,s)|^2 \, ds \, dt < \infty$$

entonces

$$(3) \qquad ||K(x(\cdot))|| \leq N||x(\cdot)||$$

que implica, por la linealidad de K, que K es continua.

No es difícil probar, como puede verse en [2], que la continuidad de una transformación lineal $T:H_1 \rightarrow H_2$ es equivalente a la existencia de una $N \geq 0$ tal que $||Tx|| \leq N||x||$, para toda x en H_1.

<u>Definición 1</u>. A una transformación lineal $T:H_1 \rightarrow H_2$ le llamaremos <u>continua</u> si existe una constante $N \geq 0$ tal que $||Tx|| \leq N||x||$ para toda x en H_1.

Geométricamente una transformación lineal $T:H_1 \rightarrow H_2$ es continua si y sólo si manda conjuntos acotados en conjuntos acotados.

<u>Definición 2.</u> Sea $T:H_1 \rightarrow H_2$ una transformación lineal y continua, definimos la norma de T, que denotaremos $||T||$, por

(4) $\quad ||T|| = \inf\{N\geq 0 \mid ||Tx||\leq N||x||,$ para toda x en $H_1\}$.

Se puede demostrar fácilmente por cálculos directos que si

(5) $$N_0 = \sup_{||x||=1} ||Tx||$$

entonces $||T||=N_0$, esto es, (4) y (5) son definiciones equivalentes de la norma de T.

Si $L(H_1,H_2)$ denota al conjunto de todas las transformaciones lineales y continuas $T:H_1\rightarrow H_2$ entonces se puede demostrar como puede verse en [5] que $L(H_1,H_2)$ es un espacio vectorial normado y completo con respecto a la norma dada en la Definición 2.

Como un ejemplo más de transformaciones lineales y continuas tenemos las proyecciones ortogonales, las cuales jugarán un papel importante en el desarrollo de la inversa generalizada.

Si M es un subespacio cerrado de H entonces por el teorema de la proyección para toda x en H existe una única x_0 en M tal que $||x-x_0|| \leq ||x-y||$ para toda y en M y

(6) $$\langle x-x_0,\ y\rangle = 0$$

para toda y en M. Esto tiene dos consecuencias:

1a). Toda x en H se puede escribir de manera única como

$$x = x_0 + x_1 \qquad (7)$$

con x_0 en M y $x_1 = x - x_0$ en $M^{\perp}$. O bien, $H = M \oplus M^{\perp}$.

2a.) La transformación $P_M : H \to H$ definida por

$$P_M\, x = x_0$$

que llamaremos la <u>proyección</u> <u>ortogonal</u> <u>de</u> H <u>sobre</u> M es lineal y continua.

La linealidad se obtiene directamente de (6) y (7). Y la continuidad de que

$$||P_M\, x||^2 = ||x_0||^2 \leq ||x_0||^2 + ||x_1||^2 = ||x||^2.$$

Más aún $||P_M|| = 1$, ya que $P_M\, y = y$ para cualquier y en M.

Por último obsérvese que $N(P_M) = \{x \mid P_M\, x = \theta\} = M$ y que $Im(P_M) = \{x_0 \mid x_0 = P_M\, x\} = M$. Esto es,

$$H = N(P_M) \oplus Im(P_M).$$

(b). Transformaciones Lineales Compactas.

Se puede demostrar como puede verse en [5] que la transformación integral $K:L^2[a,b]\to L^2[a,b]$ definida por (2) tiene la propiedad de que la cerradura de la imagen bajo K de todo conjunto acotado en $L^2[a,b]$ es un conjunto compacto en $L^2[a,b]$.

Con el siguiente ejemplo se muestra que no toda transformación lineal y continua tiene la propiedad que tienen las transformaciones integrales.

Ejemplo 2. Considérese la transformación lineal y continua $I:L^2(0,2\Pi)\to L^2(0,2\Pi)$ definida por $Ix=x$, y el conjunto $\{x_k\}$ donde

$$x_k(t) = \frac{1}{\sqrt{\Pi}}\text{ sen } k\ t,$$

para $k=1,2,\ldots$. Por cálculos directos de integración se tiene que $||x_k(\cdot)||=1$ para toda k y que $\langle x_i(\cdot),x_j(\cdot)\rangle=0$ para $i\neq j$. Ahora la imagen de $\{x_k\}$ bajo I es $\{x_k\}$, el cual no tiene un punto de acumulación ya que por el teorema de Pitágoras $||x_n(\cdot)-x_m(\cdot)||^2=||x_n(\cdot)||^2+||-x_m(\cdot)||^2=2$ para cualesquiera m y n con $m\neq n$. Y por lo tanto, la "cerradura" de $\{x_k\}$ no es "compacta".

La discusión anterior nos sugiere la siguiente

Definición 3. A una transformación lineal $K:H_1 \dashrightarrow H_2$ le llamaremos compacta si a todo conjunto acotado en H_1, K lo manda un conjunto cuya cerradura es un compacto en H_2.

De aqui en lo sucesivo K denotará una transformación lineal y compacta.

Es muy fácil probar que toda transformación lineal compacta K es continua, y que si T es una transformación lineal continua entonces KT y TK son transformaciones lineales y compactas. De la última afirmación anterior y el ejemplo 2 se tiene la siguiente

Proposición 5. Si $K:H_1 \longrightarrow H_2$ es una transformación lineal compacta que tiene inversa $K^{-1}:H_1 \longrightarrow H_2$ entonces K^{-1} no es continua a menos que H_1 y H_2 sean de dimensión finita.

(c) Transformaciones Lineales Cerradas.

En la subsección anterior vimos que la transformación inversa K^{-1} de una transformación lineal y compacta no siempre es continua. Esto es, para K^{-1} no existe $N \geq 0$ tal que $||K^{-1}x|| \leq N||x||$, razón por la cual a estas transformaciones lineales se les denomina no-acotadas.

Una clase importante de transformaciones lineales que no son continuas es la siguiente

Definición 4. A una transformación lineal $A:D\to H_2$ en donde D es un subespacio de H_1 se llama cerrada si para toda sucesión $\{x_n\}$ de D convergente en H_1, a saber $x_n\to x$ en H_1, es tal que la sucesión $\{Ax_n\}$ es convergente en H_2, a saber $Ax_n\to y$, implica que x está en D y $y=Ax$.

Es bastante directo demostrar que una transformación lineal $A:D\to H_2$ es cerrada si y sólo si la gráfica de A,

$$G = \{(x,Ax)\,|\,x \text{ en } D\}$$

es un subconjunto cerrado de $H_1 \times H_2$.

Una relación importante entre las transformaciones lineales continuas y las cerradas la da el siguiente

Teorema 2. (de la Gráfica Cerrada) Sea $A:D\to H_2$ una transformación lineal cerrada. Entonces si $D=H_1$, A es continua.

Para su demostración véase [2].

4. TRANSFORMACIONES LINEALES CONTINUAS CON SUBESPACIO IMAGEN CERRADO.

Sea $T:H_1\to H_2$ una transformación lineal continua. Ahora consideremos la ecuación

$$(1) \qquad Tx = b,$$

para la cual supóngase que no hay una solución.

Por analogía al problema inconsistente $A\bar{x}=\bar{b}$ donde A es una matriz real mxn es natural calcular una "solución aproximada" x_a para (1), según el criterio de los mínimos cuadrados. Nosotros diremos que x_a es una <u>solución aproximada</u> de (1) si

$$(2) \qquad ||Tx_a-b||\leq||Tx-b||,$$

para toda x en H_1.

Por la primera parte del teorema de la proyección, tenemos que (1) tiene una solución aproximada si y sólo si la proyección ortogonal de b está en el subespacio Im(T).

Por la segunda parte del teorema de la proyección nosotros tenemos que si $\overline{Im(T)}=Im(T)$ entonces para toda b en H (1) tiene una solución aproximada x_a.

Del párrafo anterior se sugiere preguntar: si toda transformación lineal continua $T:H_1\to H_2$ tiene subespacio Im(T) cerrado en H_2. La respuesta es negativa en general. A continuación damos un ejemplo que ilustra la situación, el cual lo demostraremos en la sec. 5 del presente capítulo.

<u>Ejemplo 1</u>. El operador integral $K:L^2[a,b]\to L^2[a,b]$ definido por $K(x(\cdot)) = \int_a^b f(\cdot,s)x(s)ds$ no tiene subespacio

$\mathrm{Im}(K)$ cerrado a menos que éste sea de dimensión finita.

Como no toda transformación lineal continua $T:H_1 \to H_2$ tiene subespacio $\mathrm{Im}(T)$ cerrado se sugiere preguntar: ¿cuáles son las clases de transformaciones lineales y continuas $T:H_1 \to H_2$ que tienen subespacio $\mathrm{Im}(T)$ cerrado? Una respuesta parcial la da el siguiente

Teorema 1. La familia de las transformaciones lineales y continuas $T:H_1 \to H_2$ con subespacio $\mathrm{Im}(T)$ cerrado en H_2 contiene a las siguientes clases

(a) A las transformaciones T con H_2 de dimensión finita o con subespacio $\mathrm{Im}(T)$ de dimensión finita.

(b) A las transformaciones T acotadas inferiormente, esto es, $||Tx|| \geq k||x||$ para toda x en H_1 y $k>0$.

(c) A las transformaciones T de la forma $T=T_1 + T_2$ donde los subespacios $\mathrm{Im}(T_1)$ y $\mathrm{Im}(T_2)$ son cerrado y de dimensión finita respectivamente.

(d) A los operadores de la forma $F=\lambda I-K$ para $\lambda \neq 0$ y con $K: H_1 \to H_1$ lineal y compacto.

Demostración.

Para (a) es obvia.

Para (b), tomemos una sucesión $\{y_n\}$ $\mathrm{Im}(T)$ con $y_n \to y$.

Ya que $||Tx|| \geq k||x||$ para toda x en H y $k>0$ se tiene que T es inyectiva. De donde la sucesión $\{y_n\}$ determina completamente la sucesión $\{x_n\}$ definida por $Tx_n=y_n$. Como $||y_n-y_m||=||Tx_n-Tx_m||\geq k||x_n-x_m||$ tenemos que la sucesión $\{x_n\}$ es de Cauchy, luego existe una x en H_1 con $x_n \longrightarrow x$, que por la continuidad de T se tiene $y_n=Tx_n \longrightarrow Tx=y$.

Para (c) es directa.

Para (d). En [6] se prueba que para cualquier $\varepsilon>0$, K se puede escribir como $K=K_1+K_2$ donde $||K_1||<\varepsilon$ y el subespacio $Im(K_2)$ es de dimensión finita. Ahora tomando $\varepsilon=\lambda$ se tiene $||\lambda^{-1}K_1||<1$ y por lo tanto $\lambda I-K_1$ es invertible sobre todo H_2 como puede verse en [2]. De donde por (c) $F=(\lambda I-K_1)-K_2$ tiene subespacio imagen cerrado en H_2.

De la parte (d) del teorema anterior se sigue que la ecuación integral inconsistente

$$\lambda\, x(t)-\int_a^b f(t,s)x(s)ds=b(t) \qquad (a\leq t\leq b)$$

siempre tiene una solución aproximada $x_a(\cdot)$ en $L^2[a,b]$ en contraste con la ecuación integral

$$\int_a^b f(t,s)x(s)ds = b(t) \qquad (a \le t \le b)$$

la cual como se vió en el ejemplo 1 no siempre tiene una solución aproximada.

5. LA INVERSA GENERALIZADA

(a) La Mejor Solución Aproximada

En la sección anterior vimos que la ecuación inconsistente

$$(1) \qquad Tx = b$$

donde $T:H_1 \to H_2$ es una transformación lineal y continua no siempre tiene una solución aproximada x_α en H_1. Y que ésta existe si $b_T = P_{\overline{Im(T)}}(b)$ está en el subespacio $Im(T)$ de H_2. Ahora si este es el caso, nosotros tenemos que x_α es una solución de la ecuación (consistente)

$$(2) \qquad Tx = b_T,$$

que es la única si el subespacio

$$N(T) = \{x \text{ en } H_1 \mid Tx = \theta_2\}$$

núcleo de T consta únicamente del vector cero θ_1 de H_1, y que en cambio, si $N(T) \neq \{\theta_1\}$ se tiene que (1) tiene una infinidad de soluciones aproximadas, a saber, de la forma

(3) $$x_a = x_{ap} + x_n$$

donde $Ax_{ap}=b_T$ y x_n es un elemento cualesquiera de $N(T)$.

De nuevo por analogía con el caso de dimensión finita se sugiere asociarle a (1) la solución aproximada x_0 de menor norma que claramente es única. Esto nos lleva a introducir la siguiente

<u>Definición 1</u>. Diremos que x_0 es la <u>mejor solución aproximada</u> (m.s.a.) de (1) cuando ésta existe si, x_0 satisface

(i). $||Tx_0-b||\leq||Tx-b||$ para toda x en H_1. Y

(ii) si $||Tx_1-b||\leq||Tx-b||$ para toda x en H_1 entonces

$$||x_0||\leq||x_1||.$$

(b). <u>La Inversa Generalizada.</u>

En la sección 3 del Capítulo anterior vimos que la m.s.a. $\bar{x}_0$ del sistema lineal $A\bar{x}=\bar{b}$ donde A denota una matriz real de dimensiones mxn, está dada por $\bar{x}_0=A^+\bar{b}$ donde A^+ denota la matriz inversa generalizada de dimensiones nxm asociada a A. Lo que sugiere introducir la siguiente

<u>Definición 2.</u> A la transformación $T^{+}:D(T^{+})\dashrightarrow H_1$ donde

(i). $D(T^{+}) = \{b \text{ en } H_2 \mid (1) \text{ tiene m.s.a.}\}$

y

(ii). definida por $T^{+}b=x_0$ donde x_0 es la m.s.a. de (1), le llamaremos la transformación <u>inversa generalizada</u> asociada a la transformación lineal y continua $T:H_1\dashrightarrow H_2$.

En el caso matricial, A^{+} define una transformación lineal y continua caracterizada por ser la única solución de las ecuaciones $AX=P_{Im(A)}$ y $XA=P_{N(A)^{\perp}}$. Es natural indagar qué propiedades de A^{+} en el caso finito son válidas para T^{+}.

Antes de tratar de establecer la linealidad de T^{+} se requiere probar que $D(T^{+})$ es un subespacio de H_2. Más aún, se puede demostrar que

<u>Proposición 1.</u> $D(T^{+})$ es un subespacio denso en H .

<u>Demostración.</u> Por el teorema de la proyección tenemos que (1) tiene una solución aproximada x_a y con esto una m.s.a. si y sólo si $P_{\overline{Im(T)}}b$ está en $Im(T)$, y esto si y sólo si b está en el subespacio $Im(T)\oplus Im(T)^{\perp}$, luego entonces $D(T^{+})=Im(T)\oplus Im(T)^{\perp}$. Del teorema de la proyección tenemos que $H_2 = \overline{Im(T)}\oplus Im(T)^{\perp}$, lo que implica que $D(T^{+})$ es un subespacio denso en H_2.

Otra caracterización equivalente de $D(T^+)$ está dada por el siguiente

Teorema 1. (Tseng) La ecuación (1) tiene una solución aproximada x_a en H_1 si y sólo si existe una constante positiva β tal que

$$(4) \qquad |\langle b,y\rangle|^2 \leq \beta\langle y, T\, T^*\, y\rangle$$

para toda y en $N(TT^*)^\perp$.

En el enunciado del Teorema anterior T^* denota a la transformación $T^*:H_2 \to H_1$ definida por $T^*y=x^*$ donde x^* satisface $\langle Tx,y\rangle = \langle x,x^*\rangle$ para toda x en H_1, la cual siempre es lineal y continua como se demuestra en [5]. Las propiedades de T^* necesarias para la demostración del teorema se dan a continuación, su demostración puede verse en la referencia anteriormente citada.

Proposición 2. Si $T:H_1 \to H_2$ es una transformación lineal continua entonces

a. $Im(T)^\perp = N(T^*)$ y $\overline{Im(T)} = N(T)^\perp$.

b. $\overline{Im(T^*)} = N(T)^\perp$ y $Im(T^*)^\perp = N(T)$.

Demostración del Teorema de Tseng.

Necesidad. Del teorema de la proyección se tiene que $H_2 = \overline{\mathrm{Im}(T)} \oplus \mathrm{Im}(T)^{\perp} = \overline{\mathrm{Im}(T)} \oplus N(T^*)$. Luego b admite la descomposición $b = b_0 + b_1$ con b_0 en $\overline{\mathrm{Im}(T)}$ y b_1 en $N(T^*)$. Como existe x_a en H_1 tal que $||Tx_a - b|| \leq ||Tx - b||$ para toda x en H_1 entonces $b_0 = Tx_a$ o bien b_0 está en $\mathrm{Im}(T)$. Tomemos $\beta = ||x_a||^2$. Ahora como $N(T^*)^{\perp} = N(TT^*)^{\perp}$ se tiene que para cualquier y en $N(TT^*)^{\perp}$

$$
\begin{aligned}
|\langle b,y\rangle| &= |\langle Tx_a + b_1, y\rangle| \\
&= |\langle Tx_a, y\rangle| \\
&= |\langle x_a, T^*y\rangle| \\
&\leq ||x_a||\ ||T^*y|| \\
&= \beta^{\frac{1}{2}}\langle y, TT^*y\rangle.
\end{aligned}
$$

Suficiencia. Definimos el producto interior $(\cdot,\cdot)$ sobre $\overline{\mathrm{Im}(T)}$ dado por $(y_1, y_2) = \langle y_1, TT^*y_2\rangle$. Sea Z el espacio completación de $\overline{\mathrm{Im}(T)}$ en la métrica definida por $(\cdot,\cdot)$.

Ya que T^* es continua en la métrica $(\cdot,\cdot)$ y H_1 es completo entonces a T^* la podemos extender a una transformación lineal y continua $B: Z \to H_1$. Escribiendo a b como arriba, significativamente $b = y_0 + y_1$ con y_0 en $\overline{\mathrm{Im}(T)}$, se tiene

$$|\langle y_0, y\rangle_2|^2 = |\langle b, y\rangle_2|^2 \leq \beta\langle y, TT^*y\rangle_2 = \beta(y,y).$$

O sea la funcional $f(y)=\langle y_0,y\rangle$ es continua sobre $Im(T)$ con respecto a la métrica $(\cdot,\cdot)$. La cual por el teorema de Hann-Banach [3] la podemos extender a una funcional lineal y continua $F:Z\to R$. Por el teorema de representación de Riesz [2] tenemos que $F(z) = (z,z_0)$ para alguna z_0 en Z. De aquí que para cualquier y en $\overline{Im(T)}$

$$f(y) = (y,z_0) = \langle y,TT^*z_0\rangle = \langle y,Tx_\alpha\rangle.$$

O sea, para toda y en $\overline{Im(T)}$, $\langle y,Tx_\alpha - y_0\rangle=0$. Luego entonces $Tx_\alpha=y_0$, y por lo tanto si (4) se satisface existe una x_α tal que $||Tx_\alpha-b|| \leq ||Tx-b||$ para toda x en H_1.

Sabiendo que $D(T^+)$ es un subespacio denso en H_2 demostraremos que T^+ es lineal, para cual se requiere el siguiente

Lema 1. $Im(T^+)\subset N(T)^\perp$.

Sea $x_0 = T^+y$ y supóngase que x_0 no está en $N(T)^\perp$. Ahora como $N(T)\cap N(T)^\perp \neq \phi$ siempre se puede tomar una solución aproximada x_1 en $N(T)$ de $Tx=y$, luego existe un vector $x_n\neq\theta$ en $N(T)$ tal que $x_0=x_1+x_n$. Por el teorema de Pitágoras se tiene que

$$||x_0||^2 = ||x_1+x_n||^2 = ||x_1||^2 + ||x_n||^2 > ||x_1||^2,$$

contradiciendo el hecho $x_0=T^+y$.

<u>Proposición 3</u>. T^+ es una transformación lineal.

<u>Demostración</u>. Sean $x_{12}=T^+(\alpha y_1+\beta y_2)$, $x_{10}=T^+ y_1$ y $x_{20}=T^+y_2$. La proposición queda demostrada si probamos que

$$x_{12}= \alpha x_{10}+\beta x_{20}.$$

Demostremos primero que $\alpha x_{10}+\beta x_{20}$ es una solución aproximada de $Tx=(\alpha y_1+\beta y_2)$. En efecto del teorema de la proyección se tiene que

$$Tx_{12}=P_{\overline{Im(T)}}(\alpha y_1+\beta y_2) =\alpha P_{\overline{Im(T)}}(y_1)+\beta P_{\overline{Im(T)}}(y_2)$$

$$=\alpha Tx_{10}+\beta Tx_{20}=T(\alpha x_{10}+\beta x_{20}),$$

de donde

$$(5) \qquad T(\alpha x_{10}+\beta x_{20}) = P_{\overline{Im(T)}}(\alpha y_1+\beta y_2)$$

y

$$(6) \qquad Tx_{12}= T(\alpha x_{10}+\beta x_{20}).$$

De (5) se tiene que $\alpha x_{10}+\beta x_{20}$ es una solución aproximada de $Tx=(\alpha y_1+\beta y_2)$, y de (6) se tiene que

$$x_{12}-(\alpha x_{10}+\beta x_{20}) = \theta$$

ya que por el lema 1 $x_{12}-(\alpha x_{10}+\beta x_{20})$ está en $N(T)\cap N(T)^{\perp} = \{\theta\}$.

Con respecto a la continuidad de T^{+} se tiene el

<u>Tecrema 2.</u> T^{+} es lineal y cerrada.

<u>Demostración.</u> Sea $\{y_n\}$ una sucesión contenida en $D(T^{+})$ con $y_n \to y$ y tal que $T^{+}y_n \to x$. Considérese la sucesión $\{x_n\}$ en H_1 definida por $x_n=T^{+}y_n$, de donde se tiene que

$$y_n = Tx_n + u_n$$

con x_n en $N(T)$ y u_n en $Im(T)^{\perp}$ ya que $H_2=\overline{Im(T)}\oplus Im(T)^{\perp}$. Claramente $x_n=T^{+}y_n \to x$ y de la continuidad de T, $Tx_n \to Tx$. Consecuentemente $u_n=(y_n-Tx_n) \to (y-Tx)$ con $y-Tx$ en $Im(T)^{\perp}$ por ser $Im(T)^{\perp}$ un subespacio cerrado de H_2, lo que demuestra que y está en $D(T^{+})$ y $T^{+}y=x$.

Una importante caracterización de T^{+} la que a su vez justifica su nombre la da el siguiente

Teorema 3. T^+ es la única transformación lineal y cerrada que satisface las ecuaciones

$$(7) \qquad XT = P_{N(T)^\perp} \text{ y } TX = P_{\overline{Im(T)}}\Big|_{D(T^+)} .$$

Para su demostración se requiere antes la siguiente

Proposición 4. Si $B=T\big|_{N(T)^\perp}$ entonces

$$(8) \qquad T^+ = B^{-1} P_{\overline{Im(T)}}\Big|_{D(T^+)} .$$

Demostración. Es inmediato verificar que $B^{-1} P_{\overline{Im(T)}}\Big|_{D(T^+)}$ está bien definida y que $D(B^{-1} P_{\overline{Im(T)}}\Big|_{D(T^+)})=D(T^+)$.

Sólo resta probar que para y en $D(T^+)$ dada,

$$(9) \qquad x_0=B^{-1} P_{\overline{Im(T)}}\Big|_{D(T^+)} (y)$$

es la m.s.a. de la ecuación $Tx=y$.

El punto x_0 dado por (9) es una solución aproximada de la ecuación $Tx=y$ ya que

$$||Tx_0-y|| = ||P_{\overline{Im(T)}}\Big|_{D(T^+)}(y)-y|| \leq ||Tx-y||$$

para toda x en H_1.

El punto x_0 dado por (9) es la m.s.a. de la ecuación $Tx=y$ porque si x_a es cualquier otra solución aproximada de la ecuación $Tx=y$ entonces $x_a=x_0+x_n$ con x_n en $N(T)$, de donde por el teorema de Pitágoras

$$||x_a||^2=||x_0+x_n||^2=||x_0||^2+||x_n||^2$$
$$\geq||x_0||^2$$

ya que por definición de B, x_0 está en $N(T)^{\perp}$.

<u>Demostración del Teorema 3.</u> Que T^+ satisface las ecuaciones dadas en (7) es directo de los hechos

$$T=BP_{N(T)^{\perp}} \text{ y } T^+ = B^{-1}\ P_{\overline{Im(T)}}\Big|_{D(T^+)} .$$

La demostración de la unicidad es esencialmente la misma que el caso matricial razón por la cual aquí, la omitimos.

6. LA INVERSA GENERALIZADA CONTINUA

Si el subespacio $Im(T)$ de T es cerrado entonces por el teorema de la proyección $D(T^+)=H_2$, luego por el teorema 3 de la sección anterior y el teorema de la gráfica cerrada T^+ es <u>continua</u>. Hemos probado el siguiente

<u>Teorema 1</u>. Si $Im(T)$ es un subespacio cerrado de H_2 entonces T^+ es una transformación lineal, continua y definida en todo H_2.

La continuidad de T^+ tiene dos importantes consecuencias prácticas: una, que para toda b en H_2 la ecuación

$$(1) \qquad Tx=b$$

siempre tiene una m.s.a. x_0; y la otra, la continuidad de la m.s.a. x_0 de la ecuación (1) lo que es muy importante desde un punto de vista computacional.

Obsérvese que todos los resultados de la sección anterior son válidos para T^+ continua.

Ahora si T^+ es tal que $||T^+y|| \leq K||y||$ para toda y en $D(T^+)$ entonces podemos extender a T^+ a todo H_2 como puede verse en [2] ya que $D(T^+)$ es un subespacio denso de H_2. En otras palabras,

$$D(T^+) = Im(T)\oplus Im(T)^{\perp}=H_2.$$

Y por lo tanto, $Im(T)$ es un subespacio cerrado de H_2. Con esto hemos probado el siguiente

<u>Teorema 2</u>. Sea $T:H_1 \longrightarrow H_2$ una transformación lineal y

continua. Entonces T^+ es continua si y sólo si Im(T) es un subespacio cerrado de H_2.

En el siguiente resultado se prueba que un operador integral de Fredholm de primera clase no tiene subespacio imagen cerrado a menos que éste sea de dimensión finita.

Teorema 3. Una transformación lineal compacta $K:H_1 \to H_2$ no tiene subespacio Im(K) cerrado a menos que éste sea de dimensión finita.

Demostración. Supóngase que $K:H_1 \to H_2$ es una transformación lineal compacta y que el subespacio Im(K) es cerado en H_2. Entonces K^+ es continua y está definida en todo H_2. De aquí que la proyección $P_{Im(K)}=KK^+$ es compacta. Pero la afirmación anterior es válida solamente si el subespacio Im(K) es de dimensión finita.

7. LAS ECUACIONES NORMALES Y LA INVERSA GENERALIZADA

En esta sección vamos a calcular a T^+ para dos casos particulares de transformaciones lineales y continuas $T:H_1 \to H_2$.

(a). La Primera Ecuación Normal.

Considérese la ecuación inconsistente

(1) $$Tx = b$$

donde $T:H_1 \to H_2$ es una transformación lineal y continua.

Ahora que exista una x_a en H_1 que minimice a $||Tx-b||$ equivale a que exista una b_r en Im(T) que minimice a $||y-b||$ sobre Im(T). Por la primera parte del teorema de la proyección y_r minimiza a $||y-b||$ sobre Im(T) si y sólo si $b-b_r$ está en Im(T) $=N(T^*)$. Y por lo tanto, si x_a es tal que $Tx_a=b_r$ entonces

$$\theta_2 = T^*(b_r-b)=T^*(Tx_a-b)=T^*Tx_a-T^*b.$$

Lo anterior es la demostración del siguiente

Teorema 1. x_a es una solución aproximada de (1) si y sólo si x_a es una solución de la ecuación normal

(2) $$T^*Tx = T^*b.$$

La importancia de la ecuación normal (2) radica en que si T^*T es invertible entonces la m.s.a. x_0 de (1) es x_a a la cual a su vez viene dada por

(3) $$x_0=(T^*T)^{-1}\ T^*b$$

de donde se sigue que

(4) $$T^{+} = (T^{*}T)^{-1}\, T^{*},$$

la cual es continua si Im(T) es un subespacio cerrado de H_2.

(b). La Segunda Ecuación Normal.

Supóngase que la ecuación

(1) $$Tx = b$$

donde $T:H_1 \longrightarrow H_2$ es una transformación lineal y continua, tiene una infinidad de soluciones.

Nuestro objetivo ahora es establecer la segunda ecuación normal ya establecida en el caso finito en la sección 2 del capítulo 1. En símbolos, x_0 es la solución de menor norma de (1) si

$$x_0 = T^{*}\, y_0$$

donde y_0 es una solución de la ecuación

$$T\, T^{*}\, y = b$$

Es inmediato demostrar que si x_0 es la solución de menor norma de (1) entonces x_0 está en $N(T)^{\perp}$. Luego para poder afirmar que existe una y_0 tal que $x_0=T^*y_0$ se requiere que $Im(T^*)=N(T)^{\perp}$ y esto último se tiene si $Im(T^*)$ es un subespacio cerrado de H_1 o equivalentemente si $Im(T)$ es un subespacio cerrado de H_2 como puede verse en [3].

Ahora si $Im(T)$ es un subespacio cerrado de H_2 entonces existe una y_0 tal que $x_0=T^*y_0$ donde y_0 es una solución de la ecuación $TT^*y=b$ ya que $TT^*y_0=Tx_0=b$.

El párrafo anterior es la demostración del siguiente

Teorema 2. Si $T:H_1 \rightarrow H_2$ es una transformación lineal, continua y con subespacio $Im(T)$ cerado en H_2 entonces x_0 es la solución de menor norma de la ecuación $Tx=b$ si y sólo si

$$x_0 = T^* y_0 \tag{5}$$

donde y_0 es una solución de la ecuación

$$T T^* y = b. \tag{6}$$

La importancia del teorema radica en el hecho que x_0 viene dada por

(7) $$x_0 = T^*(TT^*)^{-1}b,$$

si TT^* es invertible.

En el siguiente resultado se da una condición de suficiencia para la invertibilidad de TT^*.

Proposición 1. Una condición suficiente para que TT^* sea invertible es la suprayectividad de la transformación T.

Demostración. La proposición se sigue de las siguientes dos expresiones que se obtienen mediante cálculos directos

$$\mathrm{Im}(T) = \mathrm{Im}(TT^*)$$

y

$$N(TT^*) = N(T^*) = \{\theta\} .$$

BIBLIOGRAFIA

1. W.J. Kammerer y M.Z. Nashed, "Iterative Methods for Best Approximate Solutions of Linear Integral Equations of the First and Second Kinds", Journal of Mathematical Analysis and Applications, No. 40, 1972.

2. R.B. Holmes, "A Course on Optimization and Best Approximation", Springer-Verlag, 1972.

3. D.G. Luenberger, "Optimization by Vector Space Methods", Wiley & Sons, 1969.

REFERENCIAS

1. B. Noble, "Applied Linear Algebra", Prentice Hall, 1969.

2. G. Helmberg, "Introduction to Spectral Theory in Hilbert Space", Wiley Interscience, 1969.

3. D.G. Luenberger, "Optimization by Vector Space Methods", Wiley, 1969.

4. G. Bachman y L. Narici, "Functional Analysis", Academic Press International Ed., 1966.

5. A.E. Taylor, "Introduction to Functional Analysis", Wiley, 1958.

6. H. Madrid de la Vega, "Cálculo Operacional en Espacios de Hilbert y Algunas de sus Aplicaciones", Tesis Profesional, Fac. de Ciencias, UNAM, 1973.

7. E.H. Moore, "A Reciprocal of the General Algebraic Matrix", Bull. Amer. Math. Soc. 26, 1920.

8. E.H. Moore, "General Analysis: Part I", Mem. Amer. Phil. Soc. I, 1935.

9. R. Penrose, "On a Generalized Inverse for Matrices", Proc. Cambridge Phil. Soc. 51, 1955.

10. R. Penrose, "On Best Approximate Solutions of Linear Matrix Equations", Proc. Cambridge Phil. Soc. 52, 1956.

11. A. Ben Israel y A. Charnes, "Contributions to the Theory of Generalized Inverses", J. SIAM 11, 1963.

12. L.D. Pyle, "A Gradient Projection Method for Solving Programming Problems using the Generalized Inverse A^+ and the Eigenvectors e(1),...,e(n) of $I-A^+A$", Preliminary Report, Statistical and Computing Laboratory, Pardue University, Indiana, 1958.

13. J.B. Rosen, "The Gradient Projection Method for Non-Linear Programming. Part I: Linear Constraints", J. Soc. Indust. Appl. Math. 8, 1960.

14. J.B. Rosen, "The Gradient Projection Method for Nonlinear Programming. Part II: Nonlinear Constraints", J. Soc. Indust. Appl. Math. 9, 1961.

15. R.E. Kalman, "Contributions to the Theory of Optimal Control", Bol. Soc. Mat. Mexicana, 1960.

16. R.E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory", Tech. Report No. 61-1 RIAS, Baltimore, 1961.

17. J.J. Florentin, "Optimal Control of Continuous Time, Markov, Stochastic Systems", J. Electronics Control 10, 1961.

18. W.S. Loud, "Generalized Inverses and Generalized Green's Functions", SIAM J. Appl. Math. 14, 1966.

19. W.S. Loud, "Some Examples of Generalized Green's Functions and Generalized Green's Matrices", SIAM, Rev. 12, 1970.

20. W.T. Reid, "Generalized Green's Matrices for Two-Point Boundary Problems", J. Soc. Indust. Appl. Math. 15, 1967.

21. W.J. Kammerer and N.Z. Nashed, "Iterative Methods for Best Approximate Solutions of Linear Integral Equations of the First and Second Kinds", J. of Math. Anal. and Appl. 40, 1972.

22. D. Showalter, "Representation and Computation of the Pseudo inverse", Proc. Amer. Math. Soc. 18, 1967.

23. W. Petryshyn, "On Generalized Inverses and on the Uniform Convergence of $(I-\beta K)^n$ with Application to Iterative Methods", J. Math. Anal. Appl. 18, 1967.