7-39

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

TEORIA BASICA DE CUADRADOS MINIMOS
(UN ENFOQUE ELEMENTAL)

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A :

REBECA ROMERO ALVAREZ

México, D. F. 1983

I N D I C E

Página

INTRODUCCION. i

CAPITULO I

Presentación del Problema

- Ajuste con el Criterio de Cuadrados Mínimos, Caso Lineal. 1
- Interpretación Geométrica del Problema de Cuadrados Mínimos Lineal. 18

CAPITULO II

Teoría Básica

- Introducción. 44
- Existencia de la Proyección. 46
- Teorema de la Proyección. 60
- Rango Completo. 64
- Rango Deficiente. 67

CAPITULO III

Descomposición Singular y Seudoinversa

- Introducción. 82
- Obtención de la Descomposición Singular de la matriz A por medio de $A A^t$ 87
- Existencia de la Descomposición Singular en General. 97

- Interpretación Geométrica de los Valores
Singulares. 108
- La Seudoinversa. 112
- Propiedades Básicas de la Seudoinversa. 121

CAPITULO IV
Teoría de Perturbación
- Introducción. 126
- Perturbación para Inversas. 139
- Perturbación de Valores Característicos. 146
- Perturbación de Valores Singulares. 160
- Continuidad y Diferenciabilidad de Seu-
doinversas. 170
- Teoría de Perturbación de Primer Orden. 179

CAPITULO V
Significado Geométrico de las Cotas de Perturbación
- Introducción. 185
- Caso No Agudo: $\| P_1 - P_2 \| = 1$ 191
- Caso Agudo : $\| P_1 - P_2 \| < 1$ 201
- Obtención de las Cotas de Perturbación. 210
- Cotas de Perturbación de la Solución. 222

BIBLIOGRAFIA. 230

INTRODUCCION

Los objetivos de esta tesis consisten en presentar al lector:

a) Los conceptos básicos del problema de ajuste con el criterio de Cuadrados Mínimos Lineal.

b) Algunos elementos de la Teoría de Perturbación.

Estos objetivos se justifican por la importancia, cada vez mayor, que ha adquirido dentro del campo de las Matemáticas Aplicadas, la técnica de Cuadrados Mínimos. Su relevancia radica en la aplicación de este criterio, como herramienta de trabajo, en las diversas ramas de la Ciencia.

Aún cuando la mayor parte de las aplicaciones pertenecen al caso no lineal, la teoría relacionada a éste descansa sobre la del caso lineal.

En lo que se refiere a la Teoría de Perturbación, se puede comprobar el reciente desarrollo que ha tenido a partir de la década de los 50; sin embargo, su presentación dista todavía de ser accesible al lector de nivel medio, de ahí nuestro interés en presentar algunos elementos de la Teoría Básica para el caso lineal así como de la Teoría de Perturbación en un plano más asequible.

Este trabajo lo desarrollamos en cinco capítulos:

En el capítulo I explicamos en qué consiste el problema y establecemos la diferencia entre el caso lineal y el no lineal. Asímismo, damos la interpretación geométrica del problema, y hacemos hincapié en la importancia que la ortogonalidad juega en su solución. Concluimos con el Método de Gram-Schmidt como una alternativa de solución.

En el capítulo II demostramos resultados básicos que fundamentan lo anteriormente expuesto así como lo que presentamos en los capítulos siguientes. Iniciamos con la prueba de la existencia de la proyección de un vector sobre un subespacio, enunciamos el conocido Teorema de la Proyección y seguimos con la demostración de la existencia de la solución del problema en los casos cuando la matriz es de rango completo y cuando la matriz es de rango deficiente. Obtenemos, además, expresiones para la solución en los dos casos.

En el capítulo III desarrollamos algunos resultados relacionados con dos conceptos de gran relevancia para la teoría de nuestro problema.

Uno de ellos es la Descomposición Singular de una matriz A. Obtenemos esta descomposición por medio de la Descomposición Espectral de AA^t, presentación que nos es de mucha utilidad para comprender la interpretación de la descomposición. Posteriormente presentamos su demostración general.

A continuación, damos una interpretación geométrica de los valores singulares.

Después, para obtener una expresión de la solución del problema, independientemente del rango de la matriz, introducimos a la Seudoinversa de una matriz que es el otro concepto importante de este capítulo. Por último, citamos algunas propiedades básicas de la seudoinversa.

En el capítulo IV estudiamos la relación que existe entre la solución del problema de Cuadrados Mínimos y la solución de ese mismo problema "perturbado". Para ello, obtenemos resultados referentes a perturbaciones en los valores característicos, los cuales nos conducen a resultados análogos para valores singulares. Estos valores singulares constituyen uno de los medios para la obtención de resultados importantes en Teoría de Perturbación.

Antes de finalizar el capítulo desarrollamos, por su gran relevancia, los conceptos de diferenciabilidad y continuidad en seudoinversas; para después concluir con la presentación de algunas cotas de perturbación para la solución del problema.

En el último capítulo nuestro interés radica en interpretar geométricamente las cotas de perturbación. Para ello, desarrollamos algunas ideas relacionadas con la diferencia entre proyecciones ortogonales sobre subespacios "cercanos". Obtenemos, una vez más, cotas de perturbación para la solución; pero esta vez, en función de los conceptos geométricos introducidos en este mismo Capítulo.

En lo referente a fuentes de información las hemos clasificado en tres grupos. El primero comprende los libros en los que se puede consultar lo relativo a conceptos básicos de Algebra Lineal. Estos son Faddeev [3], Lang [5], Máltsev [7] y Strang [11]. El segundo grupo contiene los libros que tratan acerca de la utilidad de los aspectos numéricos de la Teoría de Perturbación mencionada y son: Kato [4], Lawson [6], Rice [8] y Stewart [9]. Y, el tercer grupo lo forman tres artículos en los que se muestra el desarrollo que ha tenido la Teoría de Perturbación recientemente. Ellos son Björck [1], Davis [2] y Stewart [10].

Por último queremos señalar que, además de los objetivos mencionados, esta tesis pretende satisfacer la necesidad urgente, dentro del campo de las Matemáticas Aplicadas, de proporcionar al lector que maneje los elementos básicos del Algebra Lineal, material accesible sobre el problema de Cuadrados Mínimos Lineal.

C A P I T U L O I

PRESENTACIÓN DEL PROBLEMA

Un problema, frecuente en la práctica, consiste en ajustar una función a una colección de datos. Es decir, dadas m parejas de puntos

$$\{(t_1, y_1), (t_2, y_2), \ldots, (t_m, y_m)\}$$

y dada una función f que depende de n+1 parámetros:

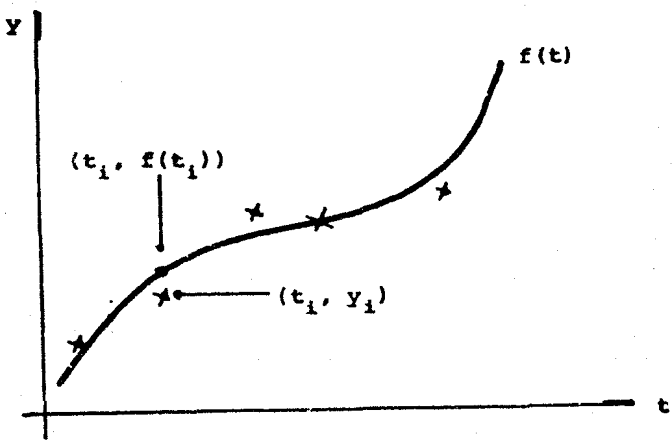
$$\xi_1, \xi_2, \ldots, \xi_n \quad \text{y de } t$$

hay que determinar aquellos valores de los parámetros para los cuales la gráfica de f pasa lo más "cerca" posible de los puntos dados.

Existen varios criterios para precisar lo que debe entenderse cuando se dice que la gráfica de f para "cerca" de los puntos dados. En este trabajo trataremos únicamente lo referente a la solución del problema de ajuste lineal con el criterio conocido con el nombre de Cuadrados Mínimos.

AJUSTE CON EL CRITERIO DE CUADRADOS MINIMOS, CASO LINEAL.

Este criterio consiste en determinar aquellos valores de los parámetros $\xi_1, \xi_2, \ldots, \xi_n$ que minimizan la suma sobre i de cuadrados de la distancia del punto (t_i, y_i) al punto $(t_i, f(t_i))$



es decir:

Encontrar los valores $\xi_1, \xi_2, \ldots, \xi_n$ que minimizan la función

$$\sum_{i=1}^{m} [y_i - f(t_i; \xi_1, \xi_2, \ldots, \xi_n)]^2$$

El cálculo de los parámetros para los cuales se alcanzan este mínimo es difícil o fácil, matemáticamente hablando, dependiendo de la forma en que éstos aparezcan en la función.

A continuación, presentaremos un ejemplo para ilustrar las dificultades que se presentan cuando se quiere determinar los valores de los parámetros en una situación concreta.

Ejemplo.

El problema consiste en determinar la curva de crecimiento de una población de bacterias. Los datos son, en este caso, el número de individuos de una especie particular de bacterias (y_i), a intervalos sucesivos de tiempo (t_i).

t_i	0	4	7.5	25	31	48.75	52	58.5	72.7	78
y_i	8	6	6	7	[illegible]	10	13	18	33	38
t_i	95	96	108	112	133	136.75	143	156.5	166.7	181
y_i	76	78	164	175	280	300	320	405	385	450

Teóricamente, se espera que el crecimiento de esta población quede bien aproximada por la curva logística, cuya expresión analítica es la siguiente:

$$f(t) = \frac{\xi_1}{1 + e^{\xi_2 - \xi_3 t}}$$

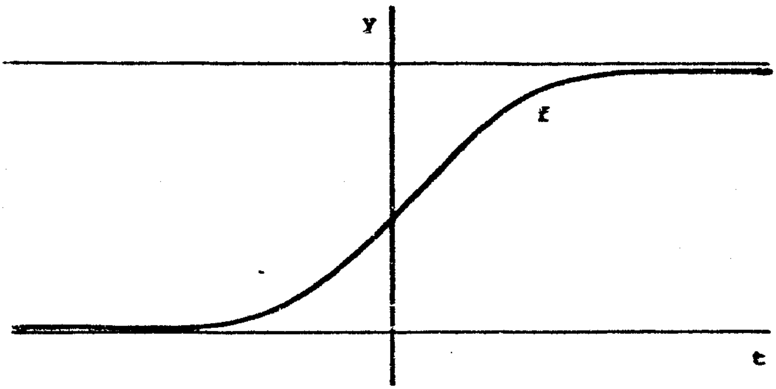
$$\text{con} \qquad \xi_3 > 0$$

en donde ξ_1, ξ_2 y ξ_3 son parámetros que dependen de la población que se trate.

Usaremos la notación

$$f(t;\ \xi_1,\ \xi_2,\ \xi_3) = \frac{\xi_1}{1 + e^{\xi_2 - \xi_3 t}}$$

para indicar que f necesita de los valores de los parámetros ξ_1, ξ_2, ξ_3 para quedar determinada. Su gráfica es de la forma indicada en la figura siguiente:



Para determinar estos parámetros se requiere minimizar la función:

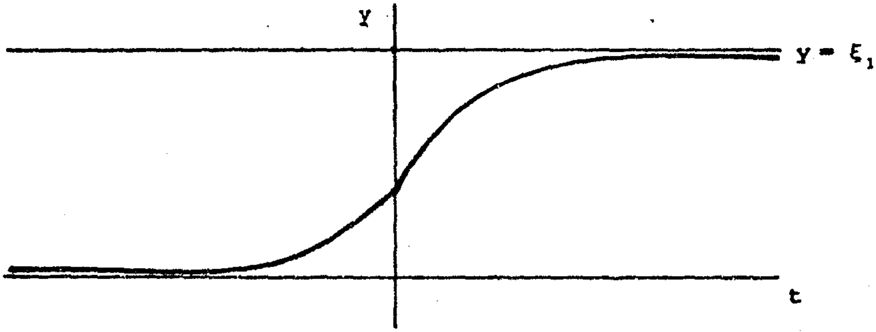
$$S(\xi_1, \xi_2, \xi_3) = \sum_{i=1}^{20} (y_i - \frac{\xi_1}{1 + e^{\xi_2 - \xi_3 t_i}})^2$$

Sin embargo, los biólogos de nuestro ejemplo, no resuelven el problema de esta forma, sino que realizan el procedimiento siguiente para encontrar los parámetros óptimos:

Para obtener el valor de ξ_1, calculan el promedio de los valores y_i correspondientes a las t_i más grandes, esto debido a que

$$y = \xi_1$$

es una asíntota de la curva buscada



Una vez calculado el valor de ξ_1 que se denotará como $\xi_1{}^*$, y que en este caso es:

$$\xi_1^* \approx \frac{450 + 385 + 405}{3}$$

$$\approx 413$$

el paso siguiente consiste en obtener estimaciones para ξ_2 y ξ_2 mediante la transformación de la función original en otra más simple en términos de ξ_2 y ξ_3.

A partir de

$$y = \frac{\xi_1{}^*}{1 + e^{\xi_2 - \xi_3 t}}$$

tenemos

$$y(1 + e^{\xi_2 - \xi_3 t}) = \xi_1{}^*$$

$$y\, e^{\xi_2 - \xi_3 t} = \xi_1{}^* - y$$

$$e^{\xi_2 - \xi_3 t} = \frac{\xi_1{}^* - y}{y}$$

de donde

$$\ln\left(\frac{\xi_1^* - y}{y}\right) = \xi_2 - \xi_3 t$$

Observemos que en esta última expresión los parámetros ξ_2 y ξ_3 aparecen en una forma muy simple y que el término que aparece en el lado izquierdo de la igualdad no representa ningún problema de cálculo, ya que el valor de ξ_1^* se ha estimado.

Entonces, si hacemos

$$g(y) = \ln\left(\frac{\xi_1^* - y}{y}\right)$$

se tiene

$$g(y) = \xi_2 - \xi_3 t$$

que es una forma transformada de la curva logística cuando ya se ha estimado el parámetro ξ_1.

Ya con esto, los biólogos de nuestro ejemplo proceden a ajustar la curva.

$$g(y) = \xi_2 - \xi_3 t$$

a los puntos $(t_i, g(y_i))$ usando el criterio de Cuadrados Mínimos, es decir:

Calculan los valores de ξ_2 y ξ_3 que minimizan la función

$$H(\xi_2, \xi_3) = \sum_{i=1}^{20} [g(y_i) - (\xi_2 - \xi_3 t_i)]^2$$

El procedimiento que utilizan es el usual para ellos y consiste en aplicar las técnicas del Cálculo Diferencial a la función H.

Se calculan las parciales de H con respecto a ξ_2 y ξ_3, se igualan a cero y se tiene lo siguiente:

$$\sum_{i=1}^{20} g(y_i) = -(\sum_{i=1}^{20} t_i)\xi_3 + 20\,\xi_2$$

$$-\sum_{i=1}^{20} g(y_i)t_i = (\sum_{i=1}^{20} t_i^2)\xi_3 - (\sum_{i=1}^{20} t_i)\xi_2$$

un sistema de ecuaciones de los más usuales, cuya solución es el mínimo buscado de la función H. Al sustituir los valores de los datos tenemos

$$\begin{aligned} 20\,\xi_2 &- 1705.4\,\xi_3 &= 26.959 \\ -1705.4\,\xi_2 &+ 205\,860.06\,\xi_3 &= 384.661 \end{aligned}$$

y la solución es

$$\xi_2 = 5.13379$$

$$\xi_3 = 0.04439$$

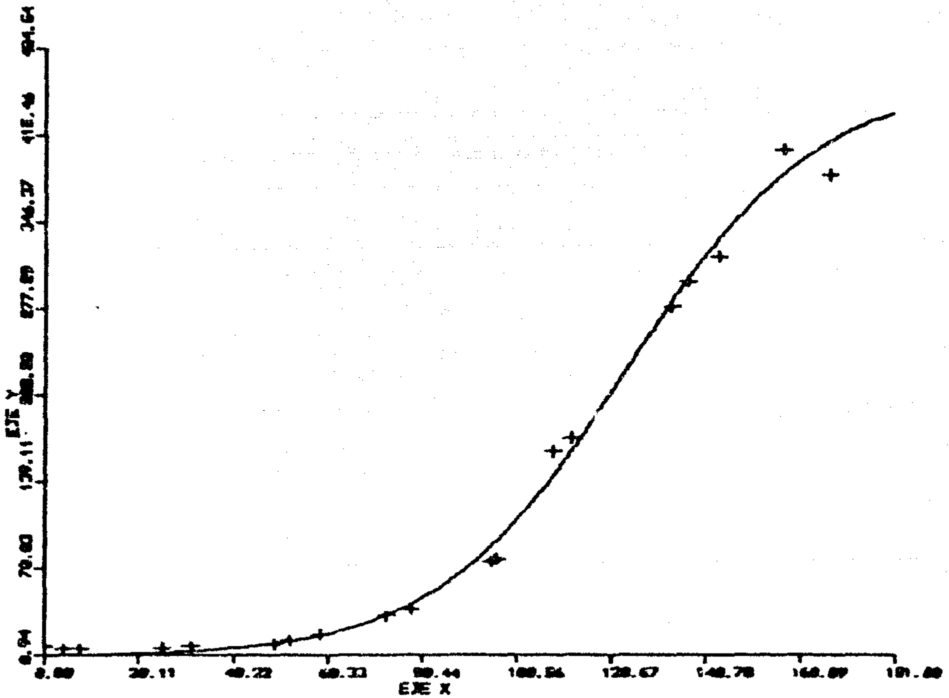
De donde la recta buscada es:

$$g(t) = 5.133 - 0.044t$$

y, por consiguiente, la curva de crecimiento deseada es:

$$f(t) = \frac{413}{1 + e^{5.133\ -.044t}}$$

y su gráfica es:



Comentarios sobre el ejemplo anterior.

Por los resultados obtenidos, podemos ver que este método parece ser muy efectivo, sin embargo, no podemos evitar hacernos la pregunta siguiente:

¿Por qué no se calcularon directamente todos los paráme tros: ξ_1, ξ_2 y ξ_3 usando el criterio de Cuadrados Mínimos?.

Para responder a esta pregunta es necesario conocer antes los problemas a los que nos tendríamos que enfrentar si lo hiciéramos.

Veamos para esto lo siguiente:

La función

$$S(\xi_1, \xi_2, \xi_3)$$

que se tendría que minimizar está dada por

$$S(\xi_1,\xi_2,\xi_3) = \sum_{i=1}^{20} (y_i - \frac{\xi_1}{1+e^{\xi_2-\xi_3 t_i}})^2$$

y para encontrar los parámetros óptimos habría que calcular las parciales de S con respecto a ξ_1, ξ_2, y ξ_3, igualarlas a cero y, posteriormente, resolver el sistema resultante que es:

$$\frac{\partial S}{\partial \xi_1} = 2 \sum_{i=1}^{20} (y_i - \frac{\xi_1}{1+e^{\xi_2-\xi_3 t}}) \cdot (\frac{-1}{1+e^{\xi_2-\xi_3 t}}) = 0$$

$$\frac{\partial S}{\partial \xi_2} = 2 \sum_{i=1}^{20} (y_i - \frac{\xi_1}{1+e^{\xi_2-\xi_3 t}}) \cdot (\frac{\xi_1 e^{\xi_2-\xi_3 t}}{(1+e^{\xi_2-\xi_3 t})^2}) = 0$$

$$\frac{\partial S}{\partial \xi_3} = 2 \sum_{i=1}^{20} (y_i - \frac{\xi_1}{1+e^{\xi_2-\xi_3 t}}) \cdot (\frac{-\xi_1 t e^{\xi_2-\xi_3 t}}{(1+e^{\xi_2-\xi_3 t})^2}) = 0$$

Es aquí donde surge el problema siguiente: No es fácil resolver este sistema de ecuaciones ya que no es de la forma lineal en ξ_1, ξ_2 y ξ_3, es decir, no es de la forma:

$$\alpha_1\xi_1 + \alpha_2\xi_2 + \alpha_3\xi_3 = \delta_1$$

$$\beta_1\xi_1 + \beta_2\xi_2 + \beta_3\xi_3 = \delta_2$$

$$\gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 = \delta_3$$

y, en general, no se conocen métodos estándar para resolver sistemas de ecuaciones simultáneas no lineales.

Por esta razón, se justifica el procedimiento especial empleado por los biólogos del ejemplo y que les permite transformar su problema a uno lineal.

Existen otros casos en donde un problema de ajuste no lineal se puede transformar en uno lineal. En este trabajo no trataremos otro ejemplo de este tipo, pues nuestro interés radicará, de aquí en adelante en comprender y desarrollar la teoría básica de aquellos problemas de ajuste que dan lugar a un sistema de ecuaciones lineales.

El desarrollo que se realizó en el ejemplo anterior sugiere, de manera natural, el deseo de investigar en qué casos se puede obtener el valor de los parámetros resolviendo un sistema de ecuaciones lineales.

Ya antes se mencionó que el criterio de Cuadrados Mínimos consiste en encontrar los valores de $\xi_1, \xi_2, \ldots, \xi_n$ que minimizan la función.

$$S(\xi_1, \xi_2, \ldots, \xi_n) = \sum_{i=1}^{m} (y_i - f(t_i; \xi_1, \xi_2, \ldots, \xi_n))^2$$

es decir

$$S(x) = \sum_{i=1}^{m} (y_i - f(t_i; x))^2$$

donde

$$x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

Ahora bien, si tomamos

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{bmatrix} \qquad y \quad F(x) = \begin{bmatrix} f(t_1; x) \\ f(t_2; x) \\ \cdot \\ \cdot \\ \cdot \\ f(t_m; x) \end{bmatrix}$$

podemos escribir a $S(x)$ en la forma

$$S(x) = \| y - F(x) \|^2$$

donde $\|\cdot\|^2$ es la norma euclideana y está dada por

$$\|a\|^2 = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2$$

si

$$a = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{bmatrix}$$

Para ver más claro el uso de esta notación, la usaremos en el ejemplo que presentamos antes.

Ejemplo:

$$x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} ; \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_{20} \end{bmatrix}$$

$$F(x) = \begin{bmatrix} \dfrac{\xi_1}{1 + e^{\xi_2 - \xi_3 t_1}} \\ \dfrac{\xi_1}{1 + e^{\xi_2 - \xi_3 t_2}} \\ \vdots \\ \dfrac{\xi_1}{1 + e^{\xi_2 - \xi_3 t_{20}}} \end{bmatrix}$$

Volviendo ahora al problema general, pasamos a calcular las parciales de S con respecto a ξ_i.

$$\frac{\partial S}{\partial \xi_k} = -2 \sum_{i=1}^{m} (y_i - f(t_i;x)) \frac{\partial f(t_i;x)}{\partial \xi_k}$$

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

Pues si el problema tiene solución, ésta debe satisfacer al sistema

$$\frac{\partial S(x)}{\partial \xi_k} = 0$$

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

Además, para que este sistema sea lineal, necesitaríamos que

$$\frac{\partial f(t_i;x)}{\partial \xi_k}$$

para $i = 1,2,\ldots,m$

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

no dependiesen de ξ_1, ξ_2 y ξ_3, es decir, necesitaríamos que fueran funciones que sólo dependieran de t_i:

$$\frac{\partial f(t_i;x)}{\partial \xi_k} = \sigma_k(t_i)$$

para $i = 1,2,\ldots,m$

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

Ahora, para saber qué condiciones debe satisfacer f para que el sistema de ecuaciones que va a determinar los parámetros.

$$\xi_1, \xi_2, \ldots, \xi_n$$

sea lineal, integramos las parciales de S son respecto a las ξ_k, para $k = 1,2,\ldots,n$; entonces tenemos:

$$f(t_i;x) = \sigma_1(t_i)\xi_1 + \sigma_2(t_i)\xi_2 + \ldots + \sigma_n(t_i)\xi_n + \theta(t_i)$$

Sin perder generalidad, supondremos que

$$\theta(t_i) \equiv 0$$

ya que no depende de los parámetros

$$\xi_1, \xi_2, \ldots, \xi_n$$

Esto, podemos resumirlo en el siguiente.

<u>Teorema</u>.

Cuando el problema de ajuste de una función con el criterio de Cuadrados Mínimos, dá lugar a la solución de un sistema de ecuaciones lineales, la función a ajustar se puede expresar de la forma siguiente:

$$f(t;\xi_1,\xi_2,\ldots,\xi_n) = \xi_1\sigma_1(t) + \xi_2\sigma_2(t) + \ldots + \xi_n\sigma_n(t)$$

donde

$$\sigma_1(t), \sigma_2(t), \ldots, \sigma_n(t)$$

son funciones de t arbitrarias, es decir, no importa la forma de las $\sigma_i(t)$.

Ahora, a partir de este resultado tenemos la siguiente:

Definición.

Un problema de ajuste de una función con el criterio de Cuadrados Mínimos se llama Lineal, si la función a ajustar se puede expresar como:

$$f(t;\xi_1,\xi_2,\ldots,\xi_n) = \xi_1\sigma_1(t)+\xi_2\sigma_2(t)+\ldots+\xi_n\sigma_n(t) \quad .$$

Este problema lo identificaremos simplemente como el de Cuadrados Mínimos Lineal y su formulación puede plantearse en términos matriciales.

En efecto, como

$$F(x) = \begin{bmatrix} f(t_1;\ x) \\ f(t_2;\ x) \\ \vdots \\ f(t_m;\ x) \end{bmatrix}$$

es decir

$$F(x) = \begin{bmatrix} \sigma_1(t_1)\xi_1+\sigma_2(t_1)\xi_2+\ldots+\sigma_n(t_1)\xi_n \\ \sigma_1(t_2)\xi_1+\sigma_2(t_2)\xi_2+\ldots+\sigma_n(t_2)\xi_n \\ \vdots \\ \sigma_1(t_m)\xi_1+\sigma_2(t_m)\xi_2+\ldots+\sigma_n(t_m)\xi_n \end{bmatrix}$$

se puede escribir como el producto de una matriz por un vector como sigue:

$$F(x) = \begin{bmatrix} \sigma_1(t_1)\sigma_2(t_1)\ldots\sigma_n(t_1) \\ \sigma_1(t_2)\sigma_2(t_2)\ldots\sigma_n(t_2) \\ \vdots \quad\quad \vdots \quad\quad \vdots \\ \sigma_1(t_m)\sigma_2(t_m)\ldots\sigma_n(t_m) \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

y entonces $$F(x) = Ax$$

donde $$A = (\sigma_k(t_i))$$

para $i = 1,2,\ldots,m$

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

Debe notarse que la matriz A es independiente de x. Y así tenemos que la nueva expresión para S(x) en términos matriciales es:

$$\begin{aligned} S(x) &= \| y - F(x) \|^2 \\ &= \| y - Ax \|^2 \end{aligned}$$

Esta expresión nos será de mucha utilidad en el transcurso de este trabajo, pues se trata de una forma compacta de denotar el problema en términos de la matriz A, de cuyas propiedades dependerá su solución.

En el problema de Cuadrados Mínimos Lineal es muy común que $m > n$, es decir, que el número de observaciones ser mayor que el de parámetros a estimar; sin embargo también se presenta el caso $m < n$.

Cuando $m > n$, como veremos más adelante, se tiene una solución única x^* si la matriz A tiene columnas independientes, hecho que proviene de que las funciones

$$\sigma_1(t),\ \sigma_2(t),\ldots,\sigma_n(t)$$

que aparecen en

$$f(t;\xi_1,\xi_2,\ldots,\xi_n) = \sigma_1(t)\xi_1 + \sigma_2(t)\xi_2+\ldots+\sigma_n(t)\xi_n$$

sean linealmente independientes, es decir, que ninguna de ellas sea combinación lineal de las restantes.

Cuando $m < n$, usualmente se pueden encontrar, muchas soluciones x tales que

$$Ax = y$$

o

$$S(x) = 0$$

En este trabajo, únicamente, trataremos el caso cuando $m > n$.

INTERPRETACION GEOMETRICA DEL PROBLEMA DE CUADRADOS MINIMOS LINEAL.

Consideremos el problema de cuadrados mínimos lineal, es decir, dados $y \in \mathbb{R}^m$ y A una matriz de $m \times n$ con $m > n$, se desea encontrar x tal que minimice la función

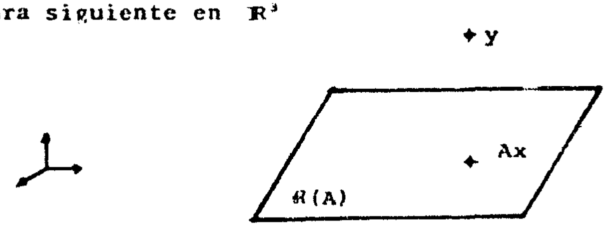
$$S(x) = \| y - Ax \|^2$$

Si consideramos a la matriz A como una transformación lineal

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

entonces la imagen de A que se denotará como $\mathcal{R}(A)$ es un subconjunto de $\mathbb{R}^m$ parecido a un plano en $\mathbb{R}^3$, y como $m > n$, usualmente el vector y no pertenece a la imagen de A.

Así que, geométricamente se tiene algo parecido a la figura siguiente en $\mathbb{R}^3$



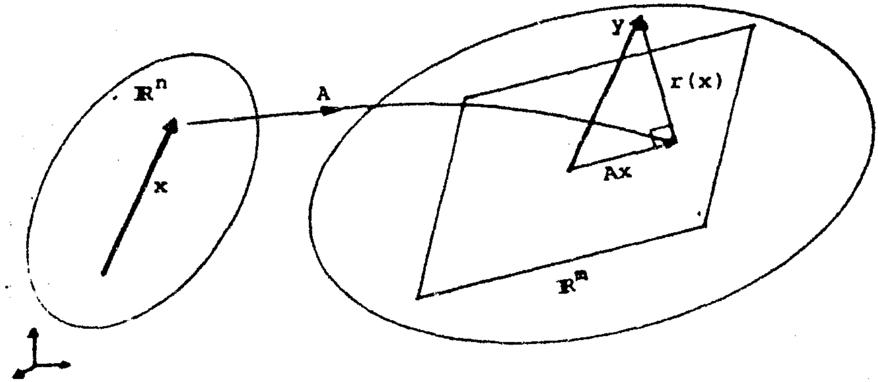
Por otro lado, para cada x, en $\mathbb{R}^n$ el vector

$$y - Ax$$

es un vector que va de un punto de $\mathcal{R}(A)$ a y, a este vector, que denotamos por

$$r(x),$$

se le conoce como el vector residual en x, véase la figura siguiente:



Como el criterio de Cuadrados Mínimos consiste en encoger x^* en $\mathbb{R}^n$ tal que

$$\| r(x^*) \|^2 < \| r(x) \|^2$$

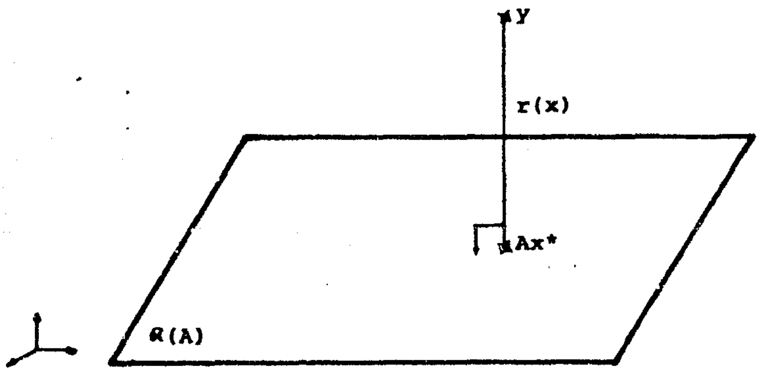
$$\text{para toda } x \text{ en } \mathbb{R}^n$$

esto, geométricamente se traduce en encontrar

$$x^* \text{ en } \mathbb{R}^n$$

tal que

Ax^* sea el punto más cercano en $\mathcal{R}(A)$ al vector y, y esto sucede, como después veremos, cuando $r(x)$ es ortogonal a $\mathcal{R}(A)$, como lo ilustra la figura que sigue:



Ortogonalidad.

En el capítulo siguiente, se demostrará que la solución al problema de Cuadrados Mínimos Lineal tiene la propiedad de que el vector

$$r(x^*) = y - Ax^*$$

es ortogonal a cada una de las columnas de A, lo que en forma compacta se puede escribir como

$$A^t(y - Ax^*) = 0$$

Por lo pronto, en esta sección sólo nos interesa hacer ver el papel tan importante que juega la ortogonalidad tanto en la teoría como en la práctica del método.

Obsérvese, que la condición de ortogonalidad

$$A^t(y - Ax^*) = 0$$

se puede escribir en la forma

$$A^tAx^* = A^t y$$

Por lo general, esta relación se utiliza para calcular x^*, se le conoce también con el nombre de "Ecuaciones Normales del problema de Cuadrados Mínimos Lineal".

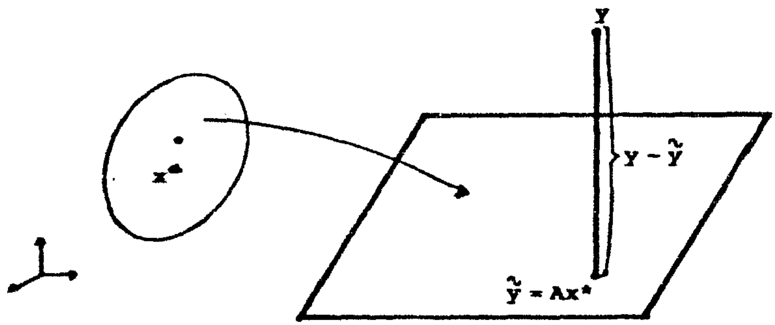
El aspecto geométrico de este planteamiento nos sugiere el siguiente procedimiento para resolver el problema de Cuadrados Mínimos Lineal:

i) Calcular el punto $\tilde{y}$ de $\mathcal{R}(A)$ más cercano a y.

ii) Resolver $Ax = \tilde{y}$.

El problema i) es conocido en la práctica como el problema de la proyección ortogonal, ya que el vector $y - \tilde{y}$ es ortogonal a todos los vectores que están en $\mathcal{R}(A)$.

Gráficamente tenemos:



Cabe, en este momento, presentar la siguiente:

Definición.

Dada $y \in \mathbb{R}^m$ y A una matriz de $m \times n$, $\tilde{y}$ en $R(A)$ se llama la proyección ortogonal de sobre $R(A)$ si el vector $y - \tilde{y}$ es ortogonal a todo vector en $R(A)$, y se denota como

$$\tilde{y} = P_{R(A)}\, y \quad .$$

El procedimiento anterior puede parecer muy complicado, pero como veremos después, no es así, ya que existen formas muy sencillas de calcular las proyecciones ortogonales.

A continuación, usando la ortogonalidad y las proyecciones ortogonales encontraremos métodos sencillos para la solución del problema de Cuadrados Mínimos.

En primer lugar, se consideraran algunos casos especiales de dicho problema para motivar los resultados generales:

1er. Caso.

Sean $A = a$ un vector de $m \times 1$, $y \in \mathbb{R}^m$ y $x = \xi$ un escalar.

$$S(\xi) = \|y - a\xi\|^2$$

es la función a minimizar.

Por otro lado, sabemos que la $\tilde{y}$ más cercana a y es tal que $y - \tilde{y}$ es ortogonal a a, es decir

$$a^t(y - \tilde{y}) = 0$$

como $\tilde{y}$ está en la imagen de $A = a$, entonces $\tilde{y} = a\tilde{\xi}$, y al sustituir tenemos

$$a^t(y - a\tilde{\xi}) = 0$$

que al multiplicar resulta

$$a^t y - a^t a\tilde{\xi} = 0$$

de donde se obtiene el valor de $\tilde{\xi}$

$$\tilde{\xi} = \frac{a^t y}{a^t a}$$

que es la solución, en este caso, al problema de Cuadrados Mínimos. Además, se puede ver que $\tilde{y}$, la proyección de y sobre a está dada por

$$\tilde{y} = a\tilde{\xi} = \frac{a^t y}{a^t a} a$$

Nótese que esta última expresión se puede escribir en la forma

$$\tilde{y} = \frac{a\, a^t}{a^t\, a} y$$

Es decir, esta última expresión la podemos ver como una fórmula para obtener directamente la proyección de y sobre a

$$\tilde{y} = P_{R(a)} y = \frac{a\, a^t}{a^t\, a} y$$

En este caso, fue muy fácil obtener la solución $\tilde{\xi}$ directamente sin calcular la proyección $\tilde{y}$. Sin embargo, como se verá más adelante, el papel de las proyecciones ortogonales en el problema de Cuadrados Mínimos Lineal es fundamental para entender la naturaleza de la solución.

2º Caso.

Sea ahora

$$A = [a_1 \mid a_2] \quad \text{de} \quad m \times 2$$

es decir, se considera el problema de Cuadrados Mínimos con dos parámetros:

$$S(\xi_1, \xi_2) = \|y - Ax\|^2$$
$$= \|y - a_1\xi_1 - a_2\xi_2\|^2$$

de acuerdo con el principio básico, la proyección $\tilde{y}$ sobre $\mathcal{R}(A)$ tiene que ser ortogonal a las columnas de A, en este caso a a_1 y a_2

Entonces tenemos

$$a_1^t\ (y - \tilde{y}) = 0$$

$$a_2^t\ (y - \tilde{y}) = 0$$

aquí se debe usar que $\tilde{y}$ está en la imagen de A, es decir en $\mathcal{R}(A)$.

Por lo tanto

$$\tilde{y} = a_1\tilde{\xi}_1 + a_2\tilde{\xi}_2 = A\tilde{x}.$$

Al sustituir esta expresión en las dos ecuaciones anteriores tenemos:

$$a_1^t(y - a_1\tilde{\xi}_1 - a_2\tilde{\xi}_2) = 0$$

$$a_2^t(y - a\ \tilde{\xi}_1 - a_2\tilde{\xi}_2) = 0.$$

Si se reescriben estas ecuaciones como un sistema obtenemos:

$$a_1^t\ a_1\xi_1 + a_1^t\ a_2\xi_2 = a_1^t\ y$$

$$a_2^t\ a_1\xi_1 + a_2^t\ a_2\xi_2 = a_2^t\ y$$

que es un sistema fácil de resolver. Sin embargo, observemos que sería más fácil de resolver si

$$a_1^t\, a_2 = 0 = a_2^t\, a_1$$

pues en este caso, se reduciría a:

$$a_1^t\, a_1\, \xi_1 = a_1^t\, y$$

$$a_2^t\, a_2\, \xi_2 = a_2^t\, y$$

El hecho de que

$$a_1^t\, a_2 = 0 = a_2^t\, a_1$$

significa que los vectores a_1 y a_2 son ortogonales. Como podemos ver, las ecuaciones anteriores indican que la solución del problema con dos parámetros se puede obtener, en este caso, por medio de la solución de dos problemas de un solo parámetro.

En efecto, si $a_1^t\, a_2 = 0 = a_2^t\, a_1$ se tiene:

$$\tilde{\xi}_1 = \frac{a_1^t\, y}{a_1^t\, a_1}$$

$$\tilde{\xi}_2 = \frac{a_2^t\, y}{a_2^t\, a_2}$$

que es la solución del problema si a_1 y a_2 son ortogonales entre sí. Además, obsérvese que obtenemos una expresión sencilla para $\tilde{y}$, la proyección de y sobre $\mathcal{R}(A)$.

En efecto, al sustituir $\tilde{\xi}_1$, $\tilde{\xi}_2$ en $\tilde{y}$ se tiene

$$\tilde{y} = a_1 \frac{a_1^t\ y}{a_1^t\ a_1} + a^2 \frac{a_2^t\ y}{a_2^t\ a_2}$$

que puede escribirse en la forma

$$\tilde{y} = \frac{a_1\ a_1^t}{a_1^t\ a_1}\ y + \frac{a_2\ a_2^t}{a_2^t\ a_2}\ y.$$

Nótese que el primer término $\frac{a_1\ a_1^t}{a_1^t\ a_1}\ y$ es la proyección de y sobre a_1 y que el segundo $\frac{a_2\ a_2^t}{a_2^t\ a_2}$ y es la proyección de y sobre a_2.

Esto nos conduce al siguiente:

Teorema.

Si las columnas de A, a_1 y a_2 son ortogonales, entonces

$$\tilde{y} = P_{\mathcal{R}(A)}\ y$$

$$= P_{a_1} y + P_{a_2} y$$

es decir, la proyección del vector y sobre $\mathcal{R}(A)$ es la suma de las proyecciones de y a lo largo de sus columnas.

En general, las columnas de A no son ortogonales, sin embargo, si seguimos adentrándonos en las características del problema, encontraremos una forma de obtener la solución sin dificultades.

A continuación, veremos que si A es de $m \times n$, con $m > n$ y A ticne columnas ortogonales, entonces, el resultado anterior sigue siendo válido.

3er. Caso.

Sea A de $m \times n$ con columnas ortogonales entre sí $a_1, a_2, \ldots, a_n$, es decir

$$A = [a_1 \mid a_2 \mid \ldots \mid a_n]$$

con $\qquad a_i^t a_j = 0 \qquad$ si $\quad i \neq j$

Para no confundir este caso especial con el caso general, cuando las columnas de A no son ortogonales, llamaremos Q a la matriz dc los n vectores ortogonales, y a sus vectores columna correspondientes

$$q_1, q_2, \ldots, q_n.$$

Así tenemos entonces en este caso

$$Q = [q_1 \mid q_2 \mid \ldots \mid q_n]$$

donde

$$q_i^t q_j = 0 \qquad \text{si} \quad i \neq j$$

La función a minimizar es

$$S(x) = S(\xi_1, \xi_2, \ldots, \xi_n) = \| y - Qx \|^2$$

$$= \| y - \xi_1 q_1 - \xi_2 q_2 - \ldots - \xi_n q_n \|^2$$

es decir, se desea

$$\min_{x} S(x) = \min_{x} \| y - Qx \|^2$$

Si $\tilde{x}$ es la solución a este problema, entonces $\tilde{y} = Q\tilde{x}$ es la proyección de y sobre $\mathcal{R}(Q)$ y, por lo tanto, se debe tener lo siguiente

$$\begin{aligned} q_1^t(y - \tilde{y}) &= 0 \\ q_2^t(y - \tilde{y}) &= 0 \\ &\vdots \\ q_n^t(y - \tilde{y}) &= 0 \end{aligned}$$

y al sustituir

$$\tilde{y} = Q\tilde{x} = q_1\tilde{\xi}_1 + q_2\tilde{\xi}_2 + \ldots + q_n\tilde{\xi}_n$$

en las ecuaciones anteriores obtenemos

$$\begin{aligned} q_1^t(y - q_1\tilde{\xi}_1 - q_2\tilde{\xi}_2 - \ldots - q_n\tilde{\xi}_n) &= 0 \\ q_2^t(y - q_1\tilde{\xi}_1 - q_2\tilde{\xi}_2 - \ldots - q_n\tilde{\xi}_n) &= 0 \\ &\vdots \\ q_n^t(y - q_1\tilde{\xi}_1 - q_2\tilde{\xi}_2 - \ldots - q_n\tilde{\xi}_n) &= 0 \end{aligned}$$

Pero como

$$q_1^t\, q_2 = q_1^t\, q_3 = \ldots = q_1^t\, q_n = 0$$

la primera ecuación se reduce a

$$q_1^t y - q_1^t\ q_1 \tilde{\xi}_1 = 0.$$

Análogamente la segunda ecuación se reduce a

$$q_2^t y - q_2^t\ q_2 \tilde{\xi}_2 = 0$$

Entonces, el sistema se reduce a

$$\begin{aligned} q_1^t\ q_1\ \tilde{\xi}_1 &= q_1^t\ y \\ q_2^t\ q_2\ \tilde{\xi}_2 &= q_2\ y \\ &\vdots \\ q_n^t\ q_n\ \tilde{\xi}_n &= q_n^t\ y \end{aligned}$$

de donde obtenemos la solución $\tilde{x}$

$$\begin{aligned} \tilde{\xi}_1 &= \frac{q_1^t\ y}{q_1^t\ q_1} \\ \tilde{\xi}_2 &= \frac{q_2^t\ y}{q_2^t\ q_2} \\ &\vdots \\ \tilde{\xi}_n &= \frac{q_n^t\ y}{q_n^t\ q_n} \end{aligned}$$

y al sustituir este valor de $\tilde{x}$ se obtiene también el valor de $\tilde{y}$ que es

$$\tilde{y} = q_1\left(\frac{q_1^t\, y}{q_1^t\, q_1}\right) + q_2\left(\frac{q_2^t\, y}{q_2^t\, q_2}\right)+\ldots+q_n\left(\frac{q_n^t\, y}{q_n^t\, q_n}\right)$$

el cual se puede reescribir como

$$\tilde{y} = \frac{q_1\, q_1^t}{q_1^t\, q_1}y + \frac{q_2\, q_2^t}{q_2^t\, q_2}y+\ldots+ \frac{q_n\, q_n^t}{q_n^t\, q_n}\, y.$$

Obsérvese que cada término es la proyección sobre una columna de Q, es decir, tenemos:

$$\tilde{y} = P_{\mathcal{R}(Q)}y = P_{q_1}y + P_{q_2}y+\ldots+P_{q_n}y$$

Esta expresión es conocida como:

EL TEOREMA DE LA PROYECCION.

La proyección ortogonal sobre $\mathcal{R}(Q)$ es igual a la suma de las proyecciones sobre sus columnas siempre y cuando dichas columnas sean ortgonales.

Antes de terminar con este caso es conveniente hacer notar lo siguiente:

La solución $\tilde{x}$ se puede expresar en una forma muy compacta:

En efecto

$$q_1^t\, q_1\, \tilde{\xi}_1 = q_1^t\, y$$

$$q_2^t\, q_2\, \tilde{\xi}_2 = q_2^t\, y$$

$$\vdots$$

$$q_n^t\, q_n\, \tilde{\xi}_n = q_n^t\, y$$

se puede escribir en términos matriciales sencillos si se introduce la matriz diagonal

$$D = \begin{bmatrix} q_1^t\, q_1 & 0 & \cdots & & 0 & 0 \\ 0 & q_2^t\, q_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & q_3^t\, q_3 & & & \cdot \\ \cdot & \cdot & 0 & \ddots & & \cdot \\ \cdot & \cdot & \vdots & & q_{n-1}^t q_{n-1} & 0 \\ \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & 0 & q_n^t q_n \end{bmatrix}$$

tenemos entonces

$$D\tilde{x} = \begin{bmatrix} q_1^t \\ q_2^t \\ \vdots \\ q_n^t \end{bmatrix} y = Q^t\, y$$

es decir

$$D\tilde{x} = Q^t\, y$$

de donde se puede despejar a $\tilde{x}$ usando la inversa de D, que también es diagonal.

Así obtenemos

$$\tilde{x} = D^{-1}Q^t y$$

Esto indica que si las columnas de la matriz son ortogonales, la obtención de la solución $\tilde{x}$ es muy sencilla, pues basta con multiplicar a y por Q^t y dividir a cada componente de $Q^t y$ por la

$$\|q_i\|^2 = q_i^t q_i$$

correspondiente. Por último, observemos que la proyección $\tilde{y}$ está dada por

$$\begin{aligned} \tilde{y} &= Q\tilde{x} \\ &= Q D^{-1} Q^t y \end{aligned}$$

Con esto terminamos el estudio de este caso.

En este momento se hace evidente la importancia de la ortogonalidad en la solución del problema de Cuadrados Mínimos Lineal, pues como ya se vió, aporta muchas ventajas en la obtención de la solución.

Sin embargo, por lo general, las columnas de la matriz A de un problema cualquiera, no son ortogonales.

¿Qué significa esto? ¿el desarrollo anterior, no tiene utilidad práctica?...

Sí, la tiene, pues es posible obtener un conjunto de vectores ortogonales a partir de las columnas de la matriz A y expresar la solución en términos de la relación entre la matriz A y la matriz formada con los vectores ortogonales.

Para aclarar esta situación y presentar el procedimiento general, consideraremos primero dos casos particulares: cuando la matriz A tiene 2 o 3 columnas.

1 er. Caso

$$A = [a_1 | a_2]_{m\times 2}$$

donde $a_1^t a_2 \neq 0$ y a_1 y a_2 no están alineados.

Interesa ahora obtener dos vectores q_1 y q_2 ortogonales entre sí a partir de a_1 y a_2

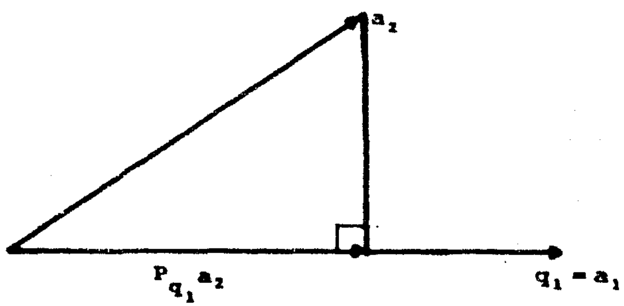
Así pues

q_1 bien puede ser el mismo a_1

entonces

$$q_1 = a_1$$

Falta ahora determinar q_2, veamos para ello la figura siguiente,



pero esto no implica ninguna dificultad, ya que calculamos la proyección de a_2 sobre q_1, que llamamos

$$\tilde{a}_2 = P_{q_1} a_2$$

y por otro lado, sabemos que $a_2 - \tilde{a}_2$ es ortogonal a q_1. Es decir informalmente podemos considerar, este modo de resolver el problema de Cuadrados Mínimos Lineal como un "generador" de vectores ortogonales.

Así definimos

$$q_1 = a_1$$

$$q_2 = a_2 - \tilde{a}_2 = a_2 - P_{q_1} a_2 .$$

Como

$$\tilde{a}_2 = P_{q_1} a_2 = \frac{q_1 \, q_1^t}{q_1^t \, q_1} \, a_2$$

$$= q_1 \frac{q_1^t \, a_2}{q_1^t \, q_1} ,$$

al sustituir se tiene:

$$q_1 = a_1$$

$$q_2 = a_2 - q_1 \frac{q_1^t \, a_2}{q_1^t \, q_1}$$

Para comprobar que q_1 y q_2 son ortogonales entre sí, calculamos $q_1^t \, q_2$.

$$q_1^t\, q_2 = q_1^t(a_2 - q_1\, \frac{q_1^t\, a_2}{q_1^t\, q_1}$$

$$= q_1^t\, a_2 - (q_1^t\, q_1)\, \frac{q_1^t\, a_2}{q_1^t\, q_1}$$

$$= q_1^t\, a_2 - q_1^t\, a_2$$

$$= 0$$

A continuación, se tratará de expresar a la matriz con columnas a_1, a_2 en términos de la matriz con columnas q_1, q_2. Para ello, despejamos a_1 y a_2 y obtenemos

$$a_1 = q_1$$

$$a_2 = q_2 + q_1\rho_{12}$$

donde

$$\rho_{12} = \frac{q_1^t\, a_2}{q_1^t\, q_1}$$

Entonces

$$A = [a_1|a_2] = [q_1|q_2 + q_1\rho_{12}]$$

$$= [q_1|q_2] \begin{bmatrix} 1 & \rho_{12} \\ 0 & 1 \end{bmatrix}$$

$$= QR$$

donde Q es una matriz $m \times 2$ y R es 2×2 . Con esto terminamos este primer caso.

2º Caso

Tenemos ahora

$$A = [\, a_1 | a_2 | a_3] ,$$

es decir A es $m \times 3$, y a_1, a_2 y a_3 no están sobre un mismo plano.

Al aplicar el mismo procedimiento anterior, cuando A era de $m \times 2$, tenemos que

$$q_1 = a_1$$

$$q_2 = a_2 - \rho_{12} q_1$$

donde

$$\rho_{12} = \frac{q_1^t \, a_2}{q_1^t \, q_1}$$

Necesitamos ahora, un vector q_3 ortogonal a los vectores q_1 y q_2. Entonces, es fácil notar que la proyección de a_3 sobre el rango de la matriz Q, cuyas columnas son q_1 y q_2, es ortogonal a dichas columnas, ya que a_3 no está en $R(Q_1)$, donde $Q_1 = [q_1 | q_2]$, debido a nuestra hipótesis de que a_1, a_2 y a_3 no están en un mismo plano.

De esta forma, el problema queda resuelto,

$$q_1 = a_1$$

$$q_2 = a_2 - q_1 \rho_{12}$$

$$q_3 = a_3 - \tilde{a}_3$$

$$= a_3 - P_{\mathcal{R}(Q_1)} a_3$$

Como

$$\tilde{a}_3 = P_{\mathcal{R}(Q_1)} a_3$$

$$= P_{q_1} a_3 + P_{q_2} a_3$$

tenemos

$$q_3 = a_3 - \frac{q_1^t\ a_3}{q_1^t\ q_1}\, q_1 - \frac{q_2^t\ a_3}{q_2^t\ q_2}\, q_2$$

y otra vez se comprueba fácilmente que q_3 es ortogonal a q_1 y q_2.

Así, para $q_1^t\ q_3$ tenemos

$$q_1^t\ q_3 = q_1^t\ a_3 - \frac{q_1^t\ a_3}{q_1^t\ q_1}\, q_1^t\ q_1 - \frac{q_2^t\ a_3}{q_2^t\ q_2}\, q_1^t\ q_2 = 0$$

ya que

$$q_1^t\ q_2 = 0$$

De aquí que

$$q_1 = a_1$$

$$q_2 = a_2 - \rho_{12} q_1$$

$$q_3 = a_3 - \rho_{23} q_2 - \rho_{13} q_1$$

donde

$$\rho_{12} = \frac{q_1^t\ a_2}{q_1^t\ q_1}\ ; \quad \rho_{13} = \frac{q_1^t\ a_3}{q_1^t\ q_1}$$

$$\rho_{23} = \frac{q_2^t\ a_3}{q_2^t\ q_2}$$

Al despejar tenemos

$$a_1 = q_1$$

$$a_2 = q_2 + q_1\ \rho_{12}$$

$$a_3 = q_3 + q_1\ \rho_{13} + q_2\ \rho_{23}$$

Y en forma matricial puede escribirse:

$$[a_1|a_2|a_3] = [q_1|q_2|q_3] \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ 0 & 1 & \rho_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

es decir

$$A = QR$$

El procedimiento para "generar" vectores ortogonales uti lizado en los dos casos puede generalizarse para una matriz A de n vectores (columnas) independientes y se conoce como el Método de Gram-Schmidt:

Si $A = [a_1|a_2|\ldots a_n]$ es una matriz de $m \times n$, entonces se puede obtener un conjunto de vectores ortogonales

$q_1, q_2, \ldots, q_n$ mediante el siguiente procedimiento:

Hacemos

$$
\begin{aligned}
q_1 &= a_1 \\
q_2 &= a_2 - \rho_{12}\, q_1 \\
q_3 &= a_3 - \rho_{13}\, q_1 - \rho_{23}\, q_2 \\
&\ \vdots \\
q_n &= a_n - \rho_{1n} q_1 - \rho_{2n} q_2 - \cdots - \rho_{n-1,n} q_{n-1}
\end{aligned}
$$

donde

$$\rho_{ij} = \frac{q_i^t\, a_j}{q_i^t\, q_i}$$

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

$$j = 2, 3, \ldots, n$$

y en forma matricial las expresiones anteriores se escriben como sigue:

$$A = [\,q_1 | q_2 | \cdots | q_n\,] \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ 0 & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \vdots & \ddots\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \rho_{n-1,n} \\ 0 & \cdots & 0 & & 1 \end{bmatrix}$$

es decir

$$A = QR$$

donde A es $m \times n$, Q es una matriz $m \times n$ con columnas ortogonales y R es una matriz $n \times n$ triangular superior con unos en la diagonal.

La propiedad que tiene Q de tener columnas ortogonales nos permite obtener la expresión:

$$Q^t Q = D = \begin{bmatrix} \|q_1\|^2 & & & 0 \\ & \|q_2\|^2 & & \\ & & \ddots & \\ 0 & & & \|q_n\|^2 \end{bmatrix}$$

Ahora bien, usando la factorización QR podemos obtener la solución del problema de cuadrados mínimos como sigue:

Si tenemos el problema

$$\min_x \|y - Ax\|^2,$$

como

$$\mathcal{R}(Q) = \mathcal{R}(A)$$

entonces $\tilde{y}$ se puede obtener directamente, en este caso, usando Q, como ya lo vimos en un ejemplo anterior:

$$\tilde{y} = QD^{-1} Q^t y$$

ahora, para obtener $\tilde{x}$, se resuelve

$$Ax = \tilde{y}$$

Y al sustituir

$$A = QR$$

y el valor de $\tilde{y}$ se obtiene

$$QR\,x = QD^{-1}\,Q^t\,y$$

Para despejar x, usaremos la siguiente propiedad de Q:

$$Q^t\,Q = D$$

Así que, al multiplicar por Q^t obtenemos

$$Q^t\,QR\,x = Q^tQD^{-1}\,Q^t\,y$$

es decir

$$DR\,x = DD^{-1}Q^t\,y,$$

de donde

$$DR\,x = Q^t\,y.$$

Entonces resulta que el sistema

$$Rx = D^{-1}\,Q^t\,y$$

es muy fácil de resolver debido a la forma triangular de R.

Nota.

Si
$$A = QR \text{ y } \tilde{D} = \begin{bmatrix} \|q_1\|^2 & & & 0 \\ & \|q_2\|^2 & & \\ & & \ddots & \\ 0 & & & \|q_n\|^2 \end{bmatrix}$$

entonces
$$\begin{aligned} A &= Q\tilde{D}\,\tilde{D}^{-1}\,R \\ &= \tilde{Q}\,\tilde{R} \end{aligned}$$

donde
$$\tilde{Q}^t\,\tilde{Q} = I$$

y
$$\tilde{R} = \begin{bmatrix} \tilde{\rho}_{11} & \tilde{\rho}_{12} & \cdots & \tilde{\rho}_{1n} \\ & \tilde{\rho}_{22} & \cdots & \tilde{\rho}_{2n} \\ & & \ddots & \vdots \\ & & & \tilde{\rho}_{nn} \end{bmatrix}$$

En este caso, la proyección de y es:
$$\tilde{y} = \tilde{Q}\,\tilde{Q}^t\,y$$

y la solución $\tilde{x}$ se obtiene a partir del sistema
$$\tilde{R}x = \tilde{Q}^t\,y.$$

Cuando las columnas de la matriz A no son linealmente independientes existen problemas: desde el punto de vista práctico, implica la reformulación del problema y desde el punto de vista matemático, el asunto lo trataremos más adelante.

C A P I T U L O II

TEORÍA BÁSICA

INTRODUCCIÓN.

En este capítulo estamos interesados en presentar las propiedades básicas de la solución del problema de ajuste lineal con el criterio de Cuadrados Mínimos.

El problema es el siguiente:

Dados A, una matriz de $m \times n$ y un vector b en $\mathbb{R}^m$, deseamos encontrar un vector x en $\mathbb{R}^n$ tal que el vector

$$r = Ax - b$$

tenga norma mínima, donde la norma de r está dada por

$$\|r\|^2 = \rho_1^2 + \rho_2^2 + \ldots + \rho_m^2$$

con

$$r = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho_m \end{bmatrix}$$

Utilizaremos la notación

$$\min_x \|b - Ax\|^2$$

para denotar el problema de ajuste lineal con el criterio de Cuadrados Mínimos, al cual, por brevedad llamaremos simplemente el Problema de Cuadrados Mínimos Lineal.

La interpretación geométrica de este problema es la siguiente:

Se busca un $\tilde{x}$ con la propiedad de que

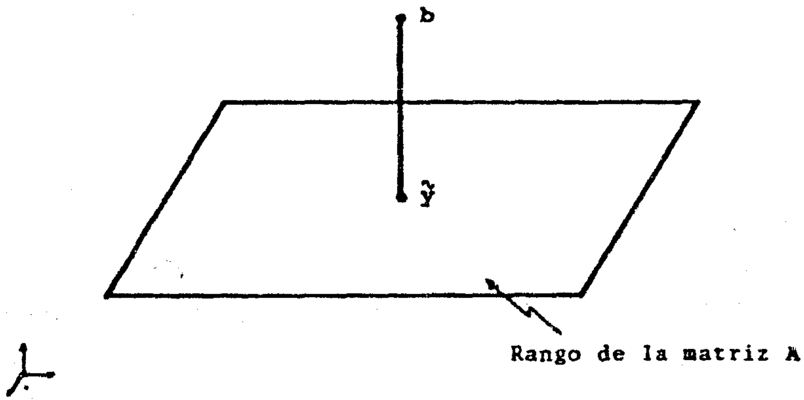
$$\tilde{y} = A\tilde{x}$$

sea el punto más cercano a b entre todas las y de la forma $y = Ax$.

El Rango de la matriz A

$$\mathcal{R}(A) = \{y \mid y = Ax\}$$

es un subespacio de $\mathbb{R}^m$, esto lo podemos ver en $\mathbb{R}^3$ como un plano que pasa por el origen. Ver la figura siguiente:



b usualmente no pertenece a $\mathcal{R}(A)$, por lo que lo podemos dibujar en la misma figura como un punto fuera del plano y entonces

$$\tilde{y} = A\tilde{x}$$

es el punto más cercano en $\mathcal{R}(A)$ a b.

Usaremos esta interpretación geométrica del problema para resolverlo en dos pasos:

i) Calcular $\tilde{y} = A\tilde{x}$, el punto en $\mathcal{R}(A)$ más cercano a y.

ii) Resolver $Ax = \tilde{y}$

A $\tilde{y}$ se le conoce con el nombre de proyección ortogonal de b sobre $\mathcal{R}(A)$.

EXISTENCIA DE LA PROYECCION.

En esta sección haremos ver que siempre existe una $\tilde{y}$ tal que

$$\min_{y \text{ en } \mathcal{R}(A)} \|b - y\|^2 = \|b - \tilde{y}\|^2$$

y que además es única.

Después veremos que la solución $\tilde{x}$ de

$$Ax = \tilde{y}$$

también existe, pero su unicidad va a depender del rango de la matriz A.

Para demostrarlo presentaremos algunos resultados elementales. La demostración de éstos se basa en las propiedades de un triángulo isósceles. El Lema siguiente establece las propiedades que nos interesan de una manera que nos será de utilidad:

Lema.

Sea y_1 y y_2, dos vectores diferentes, tales que

$$\|b - y_1\|^2 = \|b - y_2\|^2$$

$$y_M = \frac{y_1 + y_2}{2}$$

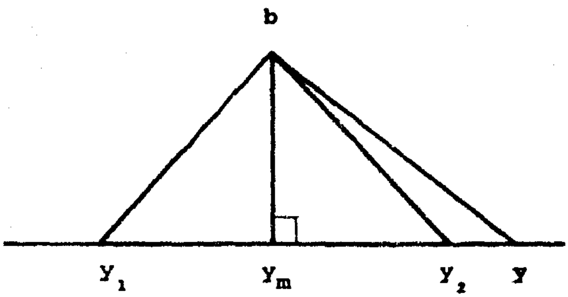
entonces

i) $(b - y_M) \perp (y_1 - y_2)$

ii) y_M es el punto más cercano a b sobre la recta que pasa por y_1 y y_2.

Demostración.

Considérese la figura siguiente donde hemos trazado la recta que contiene a y_1 y y_2. Observemos que y_M es su punto medio y que b, y_1 y y_2 forman un triángulo isósceles. Así pues, lo único que haremos ver será comprobar que y_M es su altura.



Como

$$\|b - y_1\|^2 = \|b - y_2\|^2$$

tenemos

$$(b - y_1)^t(b - y_1) = (b - y_2)^t(b - y_2)$$

$$\|y_1\|^2 - 2b^t\ y_1 = \|y_2\|^2 - 2b^t\ y_2$$

$$2b^t(y_2 - y_1) = \|y_2\|^2 - \|y_1\|^2$$

$$2b^t(y_2 - y_1) = (y_2 + y_1)^t(y_2 - y_1)$$

$$b^t(y_2 - y_1) = (\frac{y_2 + y_1}{2})^t(y_2 - y_1)$$

$$[b - \frac{(y_2 + y_1)}{2}]^t\ (y_2 - y_1) = 0$$

de donde

$$[b - \frac{(y_2 + y_1)}{2}] \perp (y_2 - y_1)$$

Además, por el Teorema de Pitágoras

$$\|b-y\|^2 = \|b-y_M\|^2 + \|y-y_M\|^2$$

de donde

$$\|b-y\|^2 > \|b-y_M\|^2$$

para toda y sobre la recta que pasa por y_1 y y_2.

En el resultado siguiente veremos que dados una recta cualquiera L y un punto P que no esté contenido en ésta, siempre existe un punto Q sobre la recta L que es el más cercano al punto P.

<u>Teorema.</u>

Sea $b \in \mathbb{R}^m$ y L una recta cualquiera en $\mathbb{R}^m$ que no contiene a b, entonces existe un punto $\tilde{y}$ sobre L que es el más cercano a b, es decir,

$$\min_{y \in L} \|b-y\|^2 = \|b-\tilde{y}\|^2$$

<u>Demostración.</u>

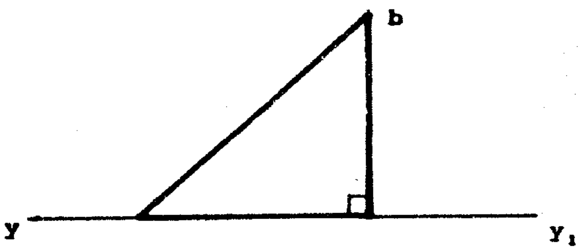
Sea $y_1 \in L$ un punto arbitrario, y sea

$$y = y_1 + ta$$

con $a \neq 0$, $t \in \mathbb{R}$, la ecuación de la recta L.

Obsérvese que si y está sobre L, y $b - y_1$ es ortogonal a L tenemos

$$\|b - y_1\|^2 + \|y_1 - y\|^2 = \|b - y\|^2$$



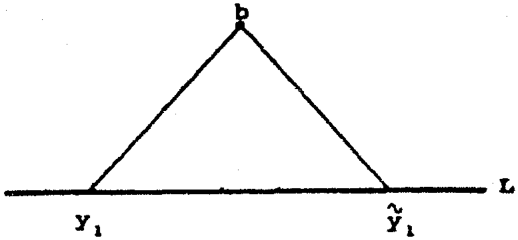
y por consiguiente

$$\|b - y_1\|^2 < \|b - y\|^2$$

para toda y sobre L. Es decir y_1 sería el punto más cercano a b, como se ve en la figura anterior.

De aquí que es natural suponer que $(b - y_1)$ no es ortogonal a L véase la siguiente figura, es decir

$$(b - y_1)^t\, a \neq 0$$



Para demostrar la afirmación del Teorema bastará con en contrar $\tilde{y}_1$ sobre L con $y_1 \neq \tilde{y}_1$ tal que

$$\|b - y_1\|^2 = \|b - \tilde{y}_1\|^2$$

si

$$\tilde{y}_1 = y_1 + \tilde{t}a$$

al sustituír tenemos

$$\|b - \tilde{y}_1\|^2 = \|b - y_1\|^2 + 2(b - y_1)^t \tilde{t}a + \tilde{t}^2 \|a\|^2$$

y resolviendo para $\tilde{t}$ se obtiene.

$$\tilde{t} = \frac{-(b - y_1)^t a}{\|a\|^2}$$

Como

$$\tilde{t} \neq 0, \quad \text{porque}$$

$$(b - y_1)^t a \neq 0$$

entonces

$$\tilde{y}_1 = y_1 + \tilde{t}a$$

está a igual distancia de b que y_1 y por el Lema anterior, su punto medio y_M es el punto más cercano a b, es decir

$$\min_{y \in L} \| b - y \|^2 = \| b - y_M \|^2 \qquad \blacksquare$$

Cuando la recta pasa por el origen las fórmulas que se obtienen son muy simples:

<u>Corolario.</u>

Si b no es un múltiplo de a, entonces el punto $\tilde{y}$ más cercano a b de la recta

$$y = \xi a$$

$$\text{con} \quad \xi \in \mathbb{R}$$

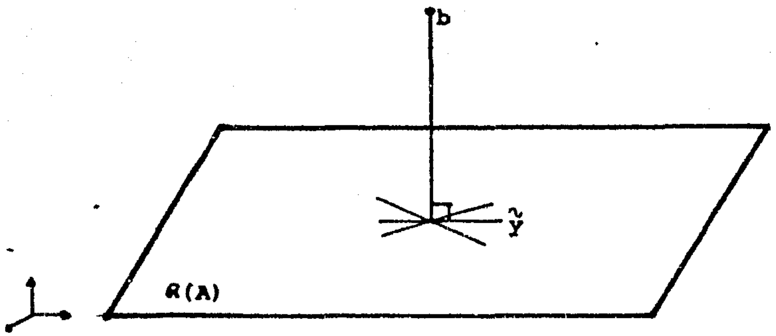
satisface

i) $b - \tilde{y}$ es ortogonal a la recta $y = \xi a$, es decir

$$(b - \tilde{y}) \perp a$$

ii) $\tilde{y} = \dfrac{a^t b}{a^t a} a = \dfrac{aa^t}{a^t a} b.$

Es natural esperar que si $\tilde{y}$ en $\mathcal{R}(A)$ es la mejor aproximación a b, entonces $(b - \tilde{y})$ es ortogonal a todas las rectas en $\mathcal{R}(A)$ que pasan por $\tilde{y}$.



Lo anterior lo establece el siguiente:

Teorema:

Si $\tilde{y} \in \mathcal{R}(A)$, es tal que

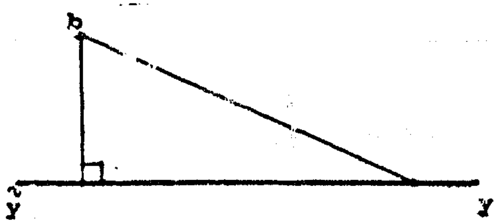
$$y \perp (b - \tilde{y})$$

para toda $y \in \mathcal{R}(A)$, entonces

$$\|b - \tilde{y}\|^2 = \min_{y \in \mathcal{R}(A)} \|b - y\|^2.$$

Demostración.

Por el Teorema de Pitágoras.



$$\|b-\tilde{y}\|^2 + \|y-\tilde{y}\|^2 = \|b-y\|^2$$

para toda $y \in \mathcal{R}(A)$

entonces

$$\|b-\tilde{y}\|^2 < \|b-y\|^2$$

para toda $y \in \mathcal{R}(A)$ ∎

A continuación probaremos en general, la existencia del punto $\tilde{y}$.

Como la demostración que presentamos sólo usa el hecho de que $\mathcal{R}(A)$ es un subespacio de $\mathbb{R}^m$, formularemos el resultado en términos de subespacios para no distraernos, por ahora, con la matriz A.

Teorema.

Si S es un subespacio de $\mathbb{R}^m$ y b es cualquier vector, entonces existe $\tilde{y}$ en S tal que

$$\min_{y \in S} \|b-y\|^2 = \|b-\tilde{y}\|^2$$

Demostración.

Sea $\{z_1, z_2, \ldots, z_k\}$ una base de S, entonces si $\tilde{y}$ existe este debe ser de la forma

$$\begin{aligned}\tilde{y} &= \alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_k z_k \\ &= Za\end{aligned}$$

donde Z es la matriz de $m \times k$ cuyas columnas son

$$z_1, z_2, \ldots, z_k$$

y a es el vector de componentes

$$\alpha_1, \alpha_2, \ldots, \alpha_k.$$

Por otro lado, si $\tilde{y}$ existe, entonces $(b - \tilde{y})$ debe ser ortogonal a todos los vectores y en S, y esto significa en particular que

$$z_i^t (b - \tilde{y}) = 0 \qquad \text{para } i = 1, 2, \ldots, k$$

es decir

$$z_i^t (b - Za) = 0 \qquad \text{para } i = 1, 2, \ldots, k$$

De donde obtenemos un sistema de ecuaciones que debe satisfacer el vector a.

Este sistema se puede escribir de la forma

$$z_i^t Za = z_i^t b \qquad \text{para } i = 1, 2, \ldots, k$$

y en forma matricial:

$$\begin{bmatrix} z_1^t\ Z \\ z_2^t\ Z \\ \cdot \\ \cdot \\ \cdot \\ z_k^t\ Z \end{bmatrix} a = \begin{bmatrix} z_1^t\ b \\ z_2^t\ b \\ \cdot \\ \cdot \\ \cdot \\ z_k^t\ b \end{bmatrix}$$

que se puede escribir en la forma

$$Z^t\ Za = Z^t\ b$$

con

$$Z = [\, z_1 \,|\, z_2 \,|\, \ldots \,|\, z_k \,]$$

Si el sistema tiene solución, ésta debe ser

$$\tilde{y} = Za$$

De aquí que todo se reduce a ver que la matriz $Z^t\ Z$ de $k \times k$ es invertible; para ello, bastará hacer ver que si

$$Z^t\ Z\tilde{a} = 0$$

entonces: $\tilde{a} = 0$

Lo anterior se obtiene al multiplicar

$$Z^t\ Z\tilde{a}$$

por $\tilde{a}^t$, ya que

$$0 = \tilde{a}^t Z^t\ Z\tilde{a} = (Z\tilde{a})^t\ (Z\tilde{a}) = \| Z\tilde{a} \|^2 = 0$$

lo que quiere decir que

$$Z\tilde{a} = 0$$

y como las columnas de Z son independientes, esto implica, a su vez, que

$$\tilde{a} = 0$$

luego entonces

$$Z^t Z \text{ es invertible}$$

y el sistema

$$(Z^t Z)a = Z^t b$$

tiene solución, y

$$\tilde{y} = Za$$

es el punto más cercano a b, pues por construcción satisface

$$Z^t(\tilde{y} - b) = 0$$

que quiere decir que $(b - \tilde{y})$ es ortogonal a la base de S, y por consiguiente a todo S.

de donde

$$\min_{y \in S} \|b - y\|^2 = \|b - \tilde{y}\|^2 \quad \blacksquare$$

Corolario.

Si $b \in \mathbb{R}^m$, entonces existe $\tilde{y}$ en $\mathcal{R}(A)$ tal que

$$\min_{y \in \mathcal{R}(A)} \| b - y \|^2 = \| b - \tilde{y} \|^2 .$$

Una vez probada la existencia de $\tilde{y}$ la demostración de que es única se sigue en forma natural.

Teorema.

Si existe $\tilde{y}$ en $\mathcal{R}(A)$ tal que

$$\min_{y \in \mathcal{R}(A)} \| b - y \|^2 = \| b - \tilde{y} \|^2 .$$

entonces $\tilde{y}$ es única.

Definición.

Dado S un subespacio de $\mathbb{R}^m$ cada b en $\mathbb{R}^m$ le podemos asociar una única s en S tal que

$$\min_{s \text{ en } S} \| b - s \|^2 = \| b - \tilde{s} \|^2$$

es decir, tenemos una función

$$P_S : \mathbb{R}^m \to \mathbb{R}^m$$

tal que

$$P_S b = \tilde{s}$$

A P_S la llamaremos la proyección ortogonal sobre S.

El siguiente resultado nos indica que la proyección ortogonal sobre un subespacio S es lineal y simétrica.

Corolario.

Si $P_S : \mathbb{R}^m \to \mathbb{R}^m$ es la proyección ortogonal sobre el subespacio S y Z es una matriz de $m \times k$ cuyas columnas son una base de S, entonces

$$P_S = Z(Z^t Z)^{-1} Z^t$$

y por consiguiente es lineal y simétrica.

Demostración.

$$\begin{aligned} P_S b &= \tilde{y} \\ &= Za \end{aligned}$$

Como se vió en la demostración del Teorema anterior

$$Z^t Za = Z^t b$$

de donde $$a = (Z^t Z)^{-1} Z^t b$$

y por lo tanto

$$P_S b = Z(Z^t Z)^{-1} Z^t b$$

TEOREMA DE LA PROYECCION.

Observemos que si b no está en el subespacio S, entonces

$$\tilde{y} = P_S\, b$$

es diferente de b y por consiguiente el vector no nulo

$$b - P_S\, b$$

es ortogonal a S.

Esta propiedad la podemos utilizar para construir bases ortogonales en cualquier subespacio S.

Teorema.

Si $\{z_1, z_2, \ldots, z_k\}$ es una base de S, y

si S_1 es el subespacio de S, cuya base es z_1,

S_2 es el subespacio de S cuya base es $\{z_1, z_2\}$

.

.

.

S_1 es el subespacio de S cuya base es $\{z_1, z_2, \ldots, z_1\}$

y también si

$$\begin{aligned} q_1 &= z_1 \\ q_2 &= z_2 - P_{S_1} z_2 \\ q_3 &= z_3 - P_{S_2} z_3 \\ &\vdots \\ q_k &= z_k - P_{S_{k-1}} z_k \end{aligned}$$

entonces

$$\{q_1, q_2, \ldots, q_k\}$$

es una base ortogonal de S.

Demostración.

Es evidente a partir de la construcción y del hecho de que

$$q_i \in S_i \quad \blacksquare$$

Observemos ahora que $\tilde{y} = P_S\, b$ tiene una expresión muy simple en la base ortogonal $\{q_1, q_2, \ldots, q_k\}$:

Supongamos que

$$\tilde{y} = \alpha_1 q_1 + \alpha_2 q_2 + \ldots + \alpha_k q_k$$

con $\alpha_i \in \mathbb{R}$

para $i = 1, 2, \ldots, k$

Entonces

$$b - \tilde{y} = b - \alpha_1 q_1 - \alpha_2 q_2 - \ldots - \alpha_k q_k$$

y como

$$(b - \tilde{y})^t\, q_i = 0$$

para $i = 1, 2, \ldots, k$

obtenemos

$$\begin{aligned} q_i^t (b - \tilde{y}) &= q_i^t b - \alpha_i q_i^t q_i \\ &= 0 \end{aligned}$$

para $i = 1, 2, \ldots, k$

de donde

$$\alpha_i = \frac{q_i^t\ b}{q_i\ q_i}$$

Al sustituir, obtenemos

$$y = \frac{q_1^t\ b}{q_1^t\ q_1}\, q_1 + \frac{q_2^t\ b}{q_2^t\ q_2}\, q_2 + \ldots + \frac{q_k^t\ b}{q_k^t\ q_k}\, q_k$$

$$= \frac{q_1\ q_1^t}{q_1^t\ q_1}\, b + \frac{q_2\ q_2^t}{q_2^t\ q_2}\, b + \ldots + \frac{q_k\ q_k^t}{q_k^t\ q_k}\, b$$

de donde

$$P_S b = \sum_{i=1}^{k} \frac{q_i\ q_i^t}{q_i^t\ q_i}\, b$$

Obsérvese que cada término de la forma

$$\frac{q_i\ q_i^t}{q_i^t\ q_i}\, b$$

es la proyección de b sobre q_i.

Lo anterior puede resumirse en el teorema siguiente:

Teorema.

Si $\{q_1, q_2, \ldots, q_k\}$ es una base ortogonal de S y P_i es la proyección ortogonal sobre q_i, entonces

$$P_S = P_1 + P_2 + \ldots + P_k$$

EXISTENCIA DE LA SOLUCION.

Hasta este momento hemos tratado acerca de la existencia y la unicidad de la proyección $\tilde{y}$.

A continuación estudiaremos la existencia y unicidad de la solución $\tilde{x}$, llamada de norma mínima del problema $\min\|y-A\tilde{x}\|^2$, así como la forma de calcularla. Esto corresponde a la segunda parte del cálculo de la Solución del Problema de Cuadrados Mínimos.

Para tratar lo anterior es necesario recordar que $\mathcal{R}(A)$ es un subespacio S, es decir

$$S = \mathcal{R}(A).$$

Presentaremos a continuación dos casos:

— Cuando las columnas de A constituyen una base de S, tendremos una sola x tal que

$$\tilde{y} = Ax$$

es decir, en este caso, existe unicidad en la representación de

$$\tilde{y} = P_S\, b$$

— Cuando las columnas de A no forman una base de S, es decir, cuando son linealmente dependientes, existirá una infinidad de representaciones

$$\tilde{y} = Ax$$

porque x no estará únicamente determinada. Y en la práctica se tiene que escoger no una $\tilde{y}$ única, sino una x única y por lo general, se elige la x de norma más pequeña.

RANGO COMPLETO.

Cuando las columnas de A forman una base para $\mathcal{R}(A)$ tenemos lo siguiente:

Teorema.

Si $\tilde{y}$ es el vector más cercano a b en $\mathcal{R}(A)$ y $\tilde{y} = A\tilde{x}$, tenemos que $\tilde{x}$ satisf. e

$$A^t(b - A\tilde{x}) = 0$$

Demostración.

Como $\quad y \perp (b - A\tilde{x})$

para toda $\quad y \in \mathcal{R}(A)$

si $\quad a_1, a_2, \ldots, a_n$

son las columnas de A, es decir, si

$$A = [a_1 | a_2 | \ldots | a_n]$$

entonces las a_i

para $i = 1,2,\ldots,n$

generan $\mathcal{R}(A)$ y por consiguiente

$$y \perp (b - A\tilde{x})$$

para toda $y \in \mathcal{R}(A)$

si y sólo si

$$a_k \perp (b - A\tilde{x})$$

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

es decir si y sólo si

$$a_k^t (b - A\tilde{x}) = 0$$

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

en forma matricial tenemos la condición buscada

$$A^t(b - A\tilde{x}) = 0. \quad \blacksquare$$

Las ecuaciones del sistema

$$A^t A\tilde{x} = A^t b$$

se conocen como las Ecuaciones Normales del problema de Cuadrados Mínimos.

Teorema.

La solución $\tilde{x}$ al problema de

$$\min_{x} \|b - Ax\|^2$$

donde A es de $m \times n$, $m > n$

puede escribirse como

$$x = (A^t A)^{-1} A^t b$$

si rango (A) = n.

Demostración.

Dado que

$$A^t Ax = A^t b$$

y como rango (A) = n, entonces A tiene columnas independientes y $A^t A$ es no singular.

Por lo tanto

$$\tilde{x} = (A^t A)^{-1} A^t b .$$

∎

Corolario.

La proyección $\tilde{y}$ de b sobre $\mathcal{R}(A)$ puede escribirse en la forma

$$\tilde{y} = A[(A^t A)^{-1} A^t b]$$

y por lo tanto

$$P_{R(A)} = A(A^t A)^{-1} A^t$$

Demostración.

Es inmediata al sustituir el valor de $\tilde{x}$ obtenido en el resultado anterior.■

Como conclusión de lo anterior tenemos el siguiente:

Teorema.

El problema de

$$\min_x \|b - Ax\|^2$$

tiene solución única si y sólo si la matriz A tiene rango completo (máximo).

RANGO DEFICIENTE.

Pasaremos ahora a tratar el caso cuando la matriz del sistema no tiene rango máximo; es decir, cuando sus columnas son linealmente dependientes.

Este caso encierra una problemática diferente, pues se tiene que elegir entre un número infinito de soluciones. Usualmente se elige la solución de norma más pequeña.

Estamos considerando ahora el problema

$$\min_{x} \| b - Ax \|^2 = \| b - \tilde{y} \|^2$$

en donde A es de $m \times n$, con $m > n$ y además rango $(A) = r < n$.

Definamos entonces el conjunto de soluciones como

$$\delta(b) = \{\tilde{x} \mid A\tilde{x} = \tilde{y} \text{ es el punto más cercano a } b\}$$

que es un conjunto de la forma

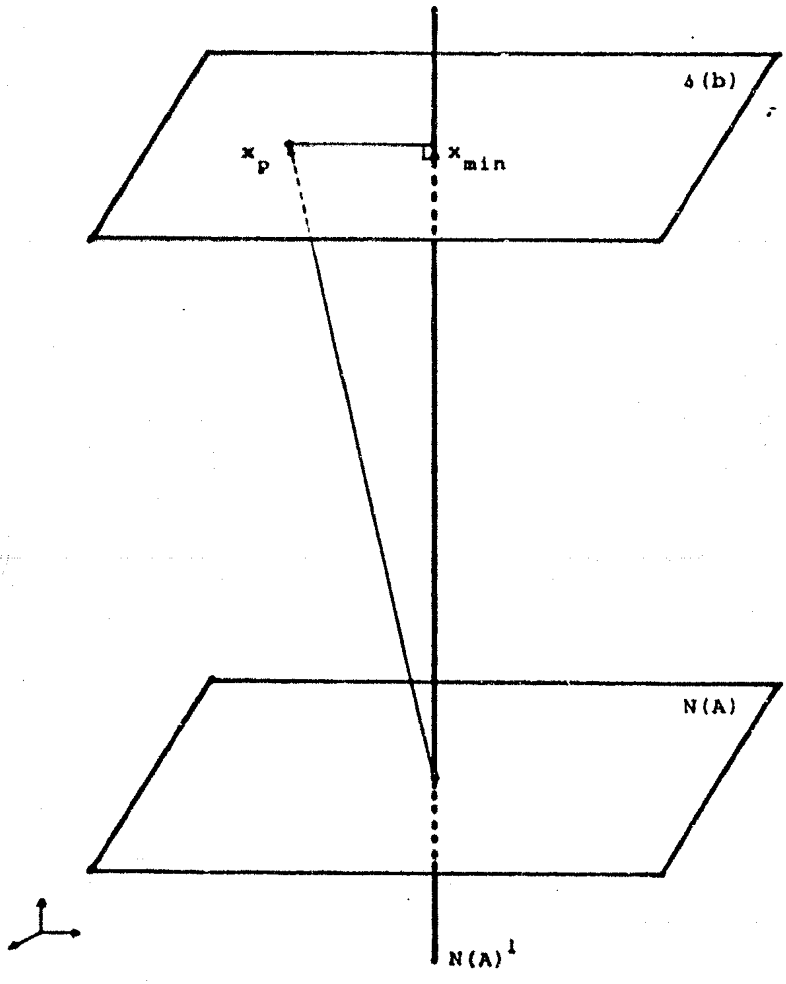
$$\begin{aligned} \delta(b) &= \tilde{x}_p + N(A) \\ &= \{\tilde{x}_p + z; \quad z \in N(A)\} \end{aligned}$$

donde

$$A\,\tilde{x}_p = \tilde{y} \qquad \text{y } N(A) \text{ es el núcleo de } A.$$

Es de particular interés el elemento x min de $\delta(b)$, que tiene norma mínima, y que además, como veremos, es fácil de calcular mediante otro problema de Cuadrados Mínimos.

Obsérvese la figura siguiente, en $\mathbb{R}^3$ en donde $\dim(N(A))=2$ es decir $N(A)$ es un plano que pasa por el origen y $\delta(b)$ es un plano paralelo a $N(A)$ que pasa por el punto x min. Notemos que $x$min es la proyección ortogonal de x_p sobre $N(A)^{\perp}$ que en la figura es una recta ortogonal a $N(A)$ que pasa por 0 y $x$min.


s(b)
x_p
x_{min}
N(A)
$N(A)^{\perp}$

<u>Teorema.</u>

Si $\tilde{y}$ es la proyección ortogonal de b sobre $\mathcal{R}(A)$ y si $\mathcal{S}(b)$ es el conjunto de x tales que

$$Ax = \tilde{y}$$

entonces $x_{min} \in \mathcal{S}(b)$, es ortogonal a $N(A)$.

<u>Demostración.</u>

Considérese el problema

$$\min_{x \in \mathcal{S}(b)} \|x\|^2$$

donde

$$\mathcal{S}(b) = \{x \mid x = x_p + z; \ z \in N(A) \ \text{y} \ Ax_p = \tilde{y}\}$$

entonces

$$\min_{x \in \mathcal{S}(b)} \|x\|^2 = \min_{z \in N(A)} \|x_p + z\|^2$$

Sea $W_{n \times k}$ tal que

$$W = [w_1, w_2, \ldots, w_k]$$

con $w_1, w_2, \ldots, w_k$ linealmente independientes y generan $N(A)$. Entonces si $z = Wh$ para alguna h, podemos escribir

$$\min_{x \in \mathcal{S}(b)} \|x\| = \min_{h} \|x_p + W_h\|^2$$

Esto como se puede ver es un problema de rango completo, de donde

$$h_{min} = -(W^t W)^{-1} W^t x_p$$

y por lo tanto

$$\begin{aligned} x_{min} &= x_p + Wh_{min} \\ &= x_p + [-W(W^t W)^{-1} W^t x_p] \\ &= [I_n - W(W^t W)^{-1} W^t] x_p \end{aligned}$$

que es una expresión para x_{min} en términos de una base para N(A) y una solución particular x_p.

Para probar que x_{min} es ortogonal a N(A) bastará con probar que x_{min} es ortogonal a las columnas de W, es decir:

$$W^t x_{min} = 0$$

En efecto

$$W^t [I - W(W^t W)^{-1} W^t] x_p =$$

$$[W^t - W^t W(W^t W)^{-1} W^t] x_p =$$

$$[W^t - W^t] x_p = 0. \quad \blacksquare$$

<u>Nota:</u>

La idea central de la demostración anterior se basa en lo siguiente:

Si $\qquad V = \{v \mid v = Ax + c;\ x \in \mathbb{R}^n\}$

el problema de resolver

$$\min_{v \in V} \| b - v \|$$

equivale al problema

$$\min_{x} \| d - Ax \|^2$$

donde

$$d = (b - c)$$

ya que

$$\min_{v \in V} \| b - v \|^2 = \min_{x} \| b - (Ax + c) \|^2$$

El siguiente paso consiste en obtener una expresión para la solución x_{min} en términos de una base de

$$N(A)^{\perp} = \mathcal{R}(A^t)$$

Esta expresión resultará más útil que la anterior pues se conoce $\mathcal{R}(A^t)$ pero no $N(A)$.

Supongamos ahora que las columnas de la matriz $n \times r$ V constituyen una base para $N(A)^{\perp}$. Entonces

$$x_{min} = Vg$$

para algún vector g

Entonces si

$$Ax_{min} = \tilde{y}$$

tenemos que

$$AVg = \tilde{y}$$

Como AV de $m \times r$, es de rango completo y además el sistema no es cuadrado, nos conviene aplicar las ecuaciones normales.

Entonces

$$g = [(AV)^t (AV)]^{-1} (AV)^t \tilde{y}$$

de donde

$$x_{min} = V [(AV)^t (AV)]^{-1} (AV)^t \tilde{y}$$

que es una expresión para la solución x_{min} en términos de una base para

$$N(A)^{\perp} = \mathcal{R}(A^t)$$

Por otro lado, cabe recordar que $z \in N(A)$ si y sólo si

$z \perp r_k$

donde r_k

para $k = 1,2,\ldots,m$

son los renglones de A. Es decir, el conjunto

$$\{r_1, r_2, \ldots, r_m\}$$

genera a $N(A)^{\perp}$.

De aquí que x_{min} puede expresarse también como

$$x_{min} = \sum_{i=1}^{m} \gamma_i r_i$$

$$= A^t g$$

para alguna g

Nota.

Como A^t no es de rango completo, A^t no puede sustituirse por V en la expresión anterior.

Ahora podemos presentar el siguiente Corolario que es básico para el resultado final.

<u>Corolario.</u>

Sea B una matriz de $r \times n$ de rango completo , y $r < n$. Entonces la solución de norma mínima x_{min}, de

$$Bx = z$$

está dada por

$$x_{min} = B^t (BB^t)^{-1} z$$

y además

$$P_{R(B^t)} = B^t (BB^t)^{-1} B$$

<u>Demostración.</u>

Como los renglones de B son linealmente independientes, constituyen una base para $N(B)^{\perp}$, entonces x_{min} se puede expresar de la forma:

$$x_{min} = B^t g$$

para alguna g

Por lo tanto

$$B(B^t g) = z$$

y como BB^t es de rango completo

$$g = (BB^t)^{-1} z$$

Al sustituir g en $x_{min} = B^t g$ se obtiene

$$x_{min} = B^t(BB^t)^{-1}z$$

y como

$$N(B)^{\perp} = \mathcal{R}(B^t) \text{ y } Bx = z$$

tenemos que

$$x_{min} = B^t(BB^t)^{-1}Bx$$

Por lo tanto

$$P_{\mathcal{R}(B^t)} = B^t(BB^t)^{-1}B \quad \blacksquare$$

Deseamos ahora obtener una expresión para la solución de norma mínima del problema de Cuadrados Mínimos, en este caso, es decir cuando rango $(A) = r < n$.

La expresión que se obtenga deberá estar como un producto de matrices de rango completo.

Para llegar a la expresión que se requiere, veamos el siguiente resultado.

<u>Teorema.</u>

Dada $A_{m \times n}$, con rango $(A) = r < n$, se puede obtener una descomposición de A como el producto de dos matrices de rango r.

$$A = QB$$

donde Q es de $m \times r$ y B de $r \times n$.

Demostración.

Por un Teorema anterior, existe una base ortogonal

$$\{q_1, q_2, \ldots, q_r\}$$

de $\mathcal{R}(A)$.

Sea $Q = [q_1 | q_2 | \ldots | q_r]$ la matriz cuyas columnas son las q_i's.

Sean también a_i

para $i = 1, 2, \ldots, n$

las columnas de A, y como $a_i \in \mathcal{R}(A)$ tenemos

$$a_i = \beta_{1i} q_i + \beta_{2i} q_2 + \ldots + \beta_{ri} q_r$$

para $i = 1, 2, \ldots, n$

Podemos escribir A en la forma siguiente. En términos de sus columnas

$$A = \sum_{i=1}^{n} a_i \; e_i^t$$

donde e_i es la base canónica de $\mathbb{R}^n$. Sustituyendo a_i en términos de los q_i, obtenemos

$$A = \sum_{i=1}^{n} \left(\sum_{j=1}^{r} \beta_{ji} \; q_j \right) e_i^t$$

$$= \sum_{j=1}^{r} q_j \left(\sum_{i=1}^{n} \beta_{ji} e_i \right)^t$$

Sea

$$b_j = \sum_{i=1}^{n} \beta_{ji} e_i$$

Entonces

$$A = \sum_{j=1}^{r} q_j b_j^t$$

Esto lo podemos escribir como el producto de Q con una matriz B cuyos renglones son las b_j, es decir :

$$A = \left(\sum_{j=1}^{r} q_j e_j^t\right)\left(\sum_{l=1}^{r} e_l b_l^t\right)$$

donde e_j y e_l son los vectores de la base canónica de $\mathbb{R}^r$,

como

$$Q = \sum_{j=1}^{r} q_j e_j^t$$

$$B = \sum_{l=1}^{r} e_l b_l^t$$

tenemos

$$A = QB$$

con $Q_{m \times r}$, $B_{r \times n}$ y como rango (A) = rango (Q) = r

tenemos que

$$\text{rango } (B) = r.$$

∎

Pasemos ahora a obtener la representación de la solución x_{min} de norma mínima, del problema

$$\min_x \|b - Ax\|^2$$

cuando A no es de rango completo.

Sea pues $A_{m \times n}$ con rango $(A) = r$. Entonces

$$A = QB$$

de donde

$$\min_x \|b - Ax\|^2 = \min_x \|b - QBx\|^2$$

y si

Bx, tenemos

$$\min_x \|b - Ax\|^2 = \min_z \|b - Qz\|^2$$

en donde Q que es de $m \times r$, tiene también rango r:

$$\text{rango } (Q) = r$$

Así pues usando las ecuaciones normales obtenemos

$$\tilde{z} = (Q^t Q)^{-1} Q^t b.$$

Además B de $r \times n$ con rango $(B) = r$ permite representar a la solución de

$$\min_{Bx = z} \|x\|^2$$

como

$$\tilde{x} = B^t (BB^t)^{-1} \tilde{z}$$

Al sustituir el valor de $\tilde{z}$ en la expresión anterior se obtiene:

$$\tilde{x} = B^t (BB^t)^{-1} (Q^t Q)^{-1} Q^t b$$

Corolario.

Sea $A\, m \times n$, con rango $(A) = r < n$, tal que

$$A = QB$$

donde Q es de $m \times r$ y B de $r \times n$

y

$$\text{rango } (Q) = \text{rango } (B) = r$$

Si B es tal que

$$B\, B^t = I_r$$

entonces la solución de

$$\min_x \| b - Ax \|^2$$

está dada por

$$\tilde{x} = B^t (Q^t\, Q)^{-1} Q^t\, b$$

y

$$P_{\mathcal{R}(A^t)} = B^t\, B.$$

Corolario.

Con las hipótesis del resultado anterior y si además Q es tal que $Q^t\, Q = I_r$ entonces la solución de

$$\min_x \| b - Ax \|^2$$

está dada por

$$\tilde{x} = B^t (BB^t)^{-1} Q^t\, b$$

y $$P_{R(A)} = QQ^t .$$

Corolario.

Con las mismas hipótesis del resultado anterior y si además B es tal que

$$BB^t = I_r$$

entonces la solución de

$$\min_x \|b - Ax\|^2$$

está dada por

$$\tilde{x} = B^t Q^t b.$$

De los resultados anteriores se deduce que si fuera posible descomponer una matriz en la forma

$$A = QB$$

donde Q tuviera columnas ortonormales y B tuviera renglones ortonormales, la expresión para la solución del problema de Cuadrados Mínimos se reduciría a una forma muy simple.

Sin embargo, esto en general no es posible, pero si en vez de ortonormalidad pedimos ortogonalidad simplemente en las columnas y renglones, veremos que sí se puede obtener la descomposición. Esto equivaldría a añadir una matriz diagonal Σ entre Q y B, es decir, buscar una descomposición

$$A = Q \Sigma B$$

En el siguiente capítulo veremos que siempre es posible obtener este tipo de descomposición.

C A P I T U L O III

DESCOMPOSICIÓN SINGULAR Y SEUDOINVERSA

INTRODUCCIÓN.

En esta sección presentaremos una serie de resultados que nos conducirán a obtener la factorización conocida como la Descomposición Singular.

Nuestro propósito es hacer ver lo siguiente:

Dada una matriz A de $m \times n$ con rango r, se puede factorizar como

$$A = U \Sigma V^t$$
$$= \sum_{i=1}^{r} \sigma_i u_i v_i^t$$

donde $U_{m\times r}$, $V_{r\times n}$ y $\Sigma_{r\times r}$ tienen la estructura siguiente:

$$U = [u_1 | u_2 | \dots | u_r]$$

$$V = [v_1 | v_2 | \dots | v_r]$$

$$\Sigma = \mathrm{diag}\,[\sigma_1, \sigma_2, \dots, \sigma_r]$$

y las σ_i's se conocen como los valores singulares de A, y las matrices U y V son tales que:

$$U^t\, U = I_r$$

$$V^t\, V = I_r$$

Además, presentaremos una de las aplicaciones más importantes de la Descomposición Singular, relacionada con la obtención de la solución de norma mínima, x_{min}, del problema de

$$\min_x \|b - Ax\|^2$$

Para comenzar, conoceremos el origen de los u_i, v_i y σ_i que aparecen en la Descomposición Singular.

Supongamos entonces que tenemos

$$A = U \Sigma V^t$$
$$= \sum_{i=1}^{r} \sigma_i u_i v_i^t$$

de donde podemos obtener los productos:

i) $$A^t A = (U \Sigma V^t)^t (U \Sigma V^t)$$
$$= V \Sigma^2 V^t$$
$$= \sum_{i=1}^{r} \sigma_i^2 v_i v_i^t$$

este primer producto indica que las $\{\sigma_i^2\}$ son los valores característicos de $A^t A$ y los v_i, los correspondientes vectores característicos ya que si multiplicamos ambos miembros de la expresión de $A^t A$ por v_i obtenemos

$$A^t A v_i = \sum_{i=1}^{r} \sigma_i^2 v_i$$

ii) $$A A^t = U \Sigma^2 U^t$$
$$= \sum_{i=1}^{r} \sigma_i^2 u_i u_i^t$$

el segundo producto nos señala que los vectores característicos de $A\ A^t$ son los u_i.

Observemos también que $\{v_i\}$, las columnas de V, forman una base ortonormal para $\mathcal{R}(A^t)$, mientras que $\{u_i\}$, las columnas de U, constituyen una base ortonormal para $\mathcal{R}(A)$.

Una vez comprendida la importancia de $A^t A$ y $A A^t$, necesitamos del Teorema siguiente para relacionar ambas matrices.

Teorema.

Si $\alpha \neq 0$, es un valor característico de $A^t A$, y u es su correspondiente vector característico, entonces α es un valor característico de $A\ A^t$ y Au es su vector característico correspondiente.

Demostración.

Tenemos

$$A^t Au = \alpha u, \text{ con } u \neq 0$$

de donde

$$Au \neq 0$$

Al multiplicar por A, se obtiene

$$(A\ A^t)Au = \alpha Au$$

■

Observemos ahora que la existencia de la factorización

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

parece indicar que si la matriz A es de rango r, también lo serán $A^t A$ y $A A^t$.

A continuación veremos que lo anterior es cierto.

Teorema.

i) $\mathcal{R}(A A^t) = \mathcal{R}(A)$

ii) $N(A) = N(A^t A)$

Demostración.

i) $\mathcal{R}(A A^t) = \mathcal{R}(A)$

Sea $z \in \mathcal{R}(A A^t) \Rightarrow z = A A^t x$

para alguna x.

$\Rightarrow z = Ay$

con $y = A^t x$

$\Rightarrow z = \mathcal{R}(A)$

por lo tanto

$$\mathcal{R}(A A^t) \subset \mathcal{R}(A)$$

Ahora bien,

sea $z \in \mathcal{R}(A) \Rightarrow z = Ay$

pero la y de norma mínima, y_{min}, es tal que

$$z = Ay_{min}$$

satisface

$$A^t g = y_{min}$$

para alguna g

como ya se vió antes

De donde

$$z = A(A^t g)$$

$$= (A A^t)g$$

$$\Rightarrow \quad z \in \mathcal{R}(A A^t)$$

y por lo tanto

$$\mathcal{R}(A) \subset \mathcal{R}(A A^t)$$

de donde

$$\mathcal{R}(A A^t) = \mathcal{R}(A). \quad \blacksquare$$

ii) $N(A) = N(A^t A)$

Sea $\quad z \in N(A) \Rightarrow Az = 0$

$$\Rightarrow A^t Az = A^t(0) = 0$$

$$\Rightarrow z \in N(A^t A)$$

por lo tanto

$$N(A) \subset N(A^t A)$$

Sea ahora

$$z \in N(A^t A) \Rightarrow A^t Az = 0,$$

de donde Az es ortogonal a los renglones de A^t, es decir, a las columnas de A, pero como $Az \in R(A)$, lo anterior sólo es posible si

$$Az = 0$$

de donde

$$z \in N(A)$$

y por lo tanto

$$N(A^t A) \subset N(A)$$

de donde

$$N(A) = N(A^t A). \blacksquare$$

Corolario.

$r = \text{rango } (A) = \text{rango } (A\, A^t) = \text{rango } (A^t A)$.

OBTENCION DE LA DESCOMPOSICION SINGULAR DE LA MATRIZ A POR MEDIO DE $A\ A^t$

Antes de presentar el desarrollo que nos conduce a la obtención de la Descomposición Singular, señalaremos las ideas principales que se utilizarán:

Trataremos de calcular la Descomposición Singular a partir de la Descomposición Espectral de $A\ A^t$.

Es conveniente hacer notar que la demostración a la que nos referimos ahora no es general, pues involucra algunas limitaciones, que oportunamente se señalarán.

Observemos, entonces que como $A\,A^t$ es una matriz simétrica, por el Teorema de la Descomposición Espectral, tenemos:

$$A\,A^t = U\,D\,U^t$$

$$= \sum_{i=1}^{r} \alpha_i\, u_i\, u_i^t$$

con

$$U = [\,u_1 | u_2 | \dots | u_r\,]$$

y $\{u_i\}$ ortonormales.

Al multiplicar por U, a la relación anterior, tenemos

$$A\,A^t\,U = U\,D$$

de donde haremos ver que $A^t U$ es "casi" la matriz V que hace falta para la factorización conocida como la Descomposición Singular.

Es decir, tendremos que obtener una V con columnas ortonormales, tal que

$$A^t\,U = V\,\bar{\Sigma}^{-1}$$

con

$$\bar{\Sigma} = \mathrm{diag}\,[\,\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_r\,]$$

para llegar a la expresión

$$AV = U\Sigma$$

que al multiplicar por V^t nos conduce al resultado deseado

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

con

$$\Sigma = \text{diag}\,[\sigma_1, \sigma_2, \ldots, \sigma_r]$$

La demostración que se obtiene al seguir estas ideas só lo es válida cuando los elementos

$$\sigma_1, \sigma_2, \ldots, \sigma_r$$

son diferentes entre sí.

Al final de la sección presentaremos una demostración general, pero en la cual no se puede interpretar el resultado en la forma tan simple como ocurre en este caso.

<u>Teorema.</u>

Sea $A_{m\times n}$ una matriz con rango $(A) = r$, $r < n$; y tal que los valores característicos de $A^t A$ no nulos, sean simples; entonces A puede factorizarse como

$$A = U_{m\times r}\, \Sigma_{r\times r}\, V^t_{r\times n}$$

en donde U y V tienen columnas ortonormales y

$$\Sigma = \text{diag}\,[\sigma_1, \sigma_2, \ldots, \sigma_r]$$

con

$$\sigma_1 > \sigma_2 > \ldots > \sigma_r$$

y que se conoce como los valores singulares de A.

Nota.

La hipótesis sobre los valores característicos de $A^t A$ no es necesaria para la demostración general.

Demostración.

Dada $A A^t$ simétrica, existe U, ortogonal y D diagonal tales que

$$A A^t = U D U^t$$
$$= \sum_{i=1}^{r} \delta_i u_i u_i^t$$

con

$$U_{mxr} = [u_1 | u_2 | \ldots | u_r]$$

y

$$D = \text{diag}\,[\delta_1, \delta_2, \ldots, \delta_r]$$

$$\text{con } \delta_i > 0$$
$$\text{para } i = 1, 2, \ldots, r$$

Multiplicando por U, obtenemos

$$A A^t U = U D$$

ya que

$$I_r = U^t U$$

Para hacer ver que las columnas de $A^t U$ son ortogonales, necesitamos probar que dichas columnas son los vectores característicos de $A^t A$.

Así que, al multiplicar por A^t, tenemos

$$A^t A(A^t U) = (A^t U)D$$

Si tomamos ahora

$$A^t U = [w_1 | w_2 | \ldots | w_r]$$

entonces

$$A^t A w_i = \delta_i w_i$$

$$\text{para } i = 1,2,\ldots,r$$

Ahora, si

$$\delta_1 > \delta_2 > \ldots > \delta_r$$

tenemos que el conjunto

$$\{w_1, w_2, \ldots, w_r\} \text{ es ortogonal}$$

Es decir, la matriz $A^t U$ tiene columnas ortogonales.

Para hacerlas ortonormales, dividimos cada columna entre su norma, y así tenemos:

$$\left[\frac{w_1}{\|w_1\|} \,\middle|\, \frac{w_2}{\|w_2\|} \,\middle|\, \cdots \,\middle|\, \frac{w_r}{\|w_r\|}\right] = [w_1|w_2|\ldots|w_r] \begin{bmatrix} \frac{1}{\|w_1\|} & & & 0 \\ & \frac{1}{\|w_2\|} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\|w_r\|} \end{bmatrix}$$

es decir

$$\left[\frac{w_1}{\|w_1\|} \,\middle|\, \frac{w_2}{\|w_2\|} \,\middle|\, \cdots \,\middle|\, \frac{w_r}{\|w_r\|}\right] = (A^t\ U)\ \overline{\Sigma}$$

con

$$\overline{\Sigma} = \operatorname{diag}\left[\frac{1}{\|w_1\|} \,\middle|\, \frac{1}{\|w_2\|} \,\middle|\, \cdots \,\middle|\, \frac{1}{\|w_r\|}\right]$$

y, regresando a la relación

$$A\ A^t\ U = U\ D$$

tenemos que al normalizarla

$$A\ A^t\ U\ \overline{\Sigma} = U\ D\ \overline{\Sigma}$$

obtenemos

$$A V = U\ \Sigma$$

$$\text{con} \quad \Sigma = D\ \overline{\Sigma}$$

y

$$V = A^t\ U\ \overline{\Sigma} = [v_1|v_2|\ldots|v_r]$$

en donde v_i son ortonormales, y además

$$\Sigma = \operatorname{diag}[\sigma_1, \sigma_2, \ldots, \sigma_r]$$

es tal que

$$\sigma_i = \frac{\delta_i}{\| w_i \|}$$

para $i = 1,2,\ldots,r$

En términos de vectores, podemos escribir

$$A\, v_i = \sigma_i\, u_i$$

para $i = 1,2,\ldots,r$.

donde

$$\{v_1, v_2, \ldots, v_r\}$$

forman una base para el $N(A)^{\perp}$.

Ahora bien, $x \in \mathbb{R}^n$, puede descomponerse en

$$x = x_1 + x_2$$

donde $x_1 \in N(A)$ y $x_2 \in N(A)^{\perp}$

y como $N(A)^{\perp}$ está generado por

$$\{v_1, v_2, \ldots, v_r\}$$

tenemos que al aplicar a x la matriz A, resulta:

$$Ax = Ax_2$$

y como

$$x_2 = P_{N(A)^\perp}\, x$$

$$= \left(\sum_{i=1}^{r} v_i\, v_i^t\right) x$$

$$= \sum_{i=1}^{r} v_i (v_i^t\, x)$$

tenemos

$$Ax = Ax_2$$

$$= A\left(\sum_{i=1}^{r} v_i\, v_i^t\right) x$$

$$= \sum_{i=1}^{r} (Av_i) v_i^t\, x$$

$$= \sum_{i=1}^{r} \sigma_i\, u_i\, v_i^t\, x$$

$$= \left(\sum_{i=1}^{r} \sigma_i\, u_i\, v_i^t\right) x$$

ya que

$$Av_i = \sigma_i\, u_i$$

$$\text{para} \quad i = 1, 2, \ldots, r$$

de donde

$$A = \sum_{i=1}^{r} \sigma_i\, u_i\, v_i^t$$

es decir

$$A = U\, \Sigma\, V^t. \qquad \blacksquare$$

Obsérvese ahora que como

$$P_{N(A)^\perp} = P_{N(A^t A)^\perp}$$

se tiene que

$$P_{N(A)^{\perp}} = \sum_{i=1}^{r} v_i \, v_i^t$$
$$= V \, V^t$$

ya que $\{v_1, v_2, \ldots, v_r\}$ generan $N(A)^{\perp}$ en el dominio de A.

También tenemos que como

$$P_{\mathcal{R}(A)} = P_{\mathcal{R}(A \, A^t)}$$
$$= \sum_{i=1}^{r} u_i \, u_i^t$$
$$= U \, U^t$$

ya que $\{u_1, u_2, \ldots, u_r\}$ generan $\mathcal{R}(A)$ en el contradominio de A.

Es conveniente hacer notar que algunas veces la Descomposicion Singular se escribe en la forma

$$A_{m\times n} = \tilde{U}_{m\times m} \, \tilde{\Sigma}_{m\times n} \, V^t_{n\times n}$$

con $\tilde{U}$ y $\tilde{V}$ ortogonales y

$$\tilde{\Sigma} = \begin{array}{cc} \left[\begin{array}{c:c} \Sigma & 0 \\ \hdashline 0 & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{array}\right] & \begin{array}{c} r \\ \\ m-r \end{array} \\ \begin{array}{cc} \quad r \quad & \quad n-r \quad \end{array} & \end{array}$$

La obtención de esta factorización implica completar el conjunto $\{v_1, v_2, \ldots, v_r\}$ para formar una base ortonormal en el dominio de A, y hacer lo mismo con

$$\{u_1, u_2, \ldots, u_r\}$$

en el contradominio.

Esto sería, formar $\tilde{V}$

$$\tilde{V} = [\, V_r \,|\, V_{1\,n-r} \,]$$

con columnas ortonormales, lo que equivaldría a obtener $\tilde{V}_{nxn}$ ortogonal y hacer lo análogo en el contradominio:

Formar

$$\tilde{U} = [\, U_r \,|\, U_{1\,m-r} \,]$$

con $\tilde{U}_{mxm}$ ortogonal.

Para así llegar a calcular

$$A\,\tilde{V} = \tilde{U}\,\tilde{\Sigma}$$

con

$$\tilde{\Sigma} = \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{array}\right]$$

y después, multiplicando por $\tilde{V}^t$ obtener

$$A = \tilde{U}_{m \times m} \tilde{\Sigma}_{m \times n} \tilde{V}^t_{n \times n}$$

EXISTENCIA DE LA DESCOMPOSICION SINGULAR EN GENERAL.

Antes de formular el resultado anterior en general y demostrarlo, recordaremos la definición de norma de una matriz*.

Para ello, consideraremos el problema de ampliar el concepto de norma euclideana de un vector a una matriz:

Definición.

Si $A_{m \times n}$, entonces la norma de A: $\|A\|$, se define como

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

La definición anterior también puede expresarse como sigue

$$\|A\| = \max_{\|x\| = 1} \|Ax\|$$

debido a la continuidad de $\|Ax\|$.

Con esta última definición podemos interpretar geométricamente a la norma de $A: \|A\|$, como la longitud del vector

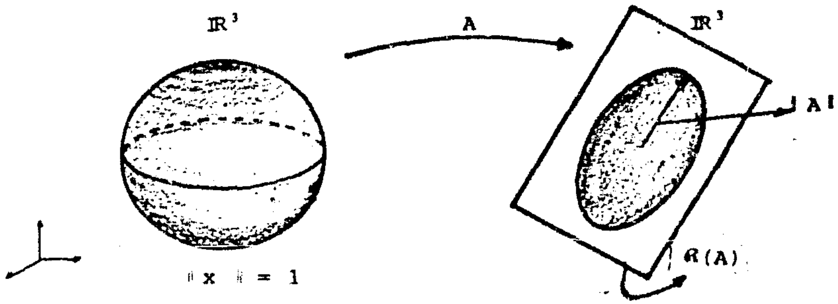
* Nos referimos siempre a la norma euclideana.

"más largo de la imagen de la "bola unitaria" bajo A.

Es decir, el vector "más grande" del conjunto

$$Z = \{z/z = Ax;\ \| x \| = 1\}$$

Si rango (A) = r, la frontera de este conjunto imagen es, de hecho, un elipsoide de r dimensiones en $\mathbb{R}^m$.



La figura anterior nos ilustra la interpretación geométrica de la norma cuando m = 3 y rango (A) = r = 2

Para presentar la demostración general sobre la existencia de la Descomposición Singular requerimos de los resultados siguientes:

<u>Lema.</u>

Existen $x \in \mathbb{R}^n$ y $y \in \mathbb{R}^m$ tales que

$$Ax_1 = \sigma_1 y_1$$

con

$$\|x_1\| = \|y_1\| = 1 \quad y \quad \sigma_1 = \|A\|.$$

Demostración.

Como

$$\|A\| = \max_{\|x\|=1} \|A_x\|$$

entonces, por continuidad, existe x_1, tal que

$$\|A\| = \|Ax_1\|$$

y

$$\|x_1\| = 1$$

Si ahora, hacemos

$$\sigma_1 = \|A\|$$

y

$$y_1 = \frac{Ax_1}{\|Ax_1\|} = \frac{Ax_1}{\|A\|}$$

$$= \frac{Ax_1}{\sigma_1}$$

entonces

$$Ax_1 = \sigma_1 \, y_1$$

con

$$\|x_1\| = \|y_1\| = 1.$$

■

Para probar la existencia de la Descomposición Singular utilizaremos la idea siguiente:

Se buscarán matrices U_1^t y V_1 ortogonales tales que

$$\begin{array}{cc} U^t_{m\times m}\, A_{m\times n}\, V_{n\times n} = \begin{bmatrix} \sigma_1 & 0 \; . \; . \; . \; 0 \\ 0 & \\ . & \\ . & B \\ . & \\ 0 & \end{bmatrix} & \begin{array}{c} 1 \\ \\ \\ m-1 \\ \\ \\ \end{array} \\ \qquad\qquad\qquad\qquad\quad 1 \qquad\quad n-1 & \end{array}$$

y después se procederá por inducción.

De acuerdo con la idea anterior, el Lema siguiente es relevante en la demostración.

<u>Lema.</u>

Dados u, v tales que

$$\|u\| = \|v\| \neq 0$$

existe P, una transformación ortogonal tal que

$$Pu = v$$

<u>Demostración.</u>

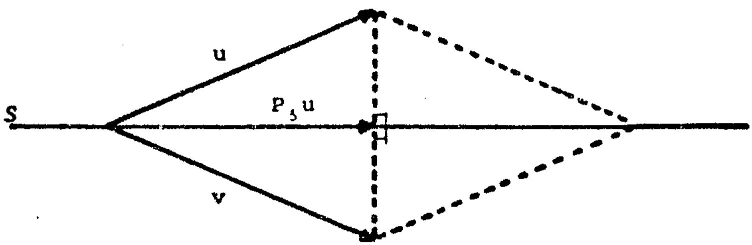
Se hará ver que P es una reflexión sobre la bisectriz de u, v, vease como ilustración la figura siguiente.

Sea entonces

$$S = \{z / z = \lambda(u+v) \; ; \; \lambda \in \mathbb{R}\}$$

y sea $\qquad z_0 \in S$ tal que $\|z_0\| = 1$

Entonces la proyección de u sobre S está dada por:



$$P_S \, u = z_0^t \, u \cdot z_0$$

como $\|z_0\| = 1$, al reordenar, obtenemos

$$P_S \, u = z_0^t \, z_0 \, u$$

entonces

$$v = u - 2(u - P_S \, u)$$

$$= u - 2(u - z_0^t \, z_0 u)$$

$$v = (2 \, z_0 \, z_0^t - I) \, u$$

de donde, al tomar a P como

$$P = 2 \, z_0 \, z_0^t - I$$

obtenemos el resultado deseado, es decir, P es tal que

$$Pu = v. \quad \bullet$$

Con estos antecedentes pasamos a la demostración general del Teorema de la Descomposición Singular.

Teorema.

Sea A una matriz de $m \times n$, entonces existen $U_{m \times m}$ ortogonal, $V_{n \times n}$ ortogonal y $\Sigma_{m \times n}$ diagonal con elementos

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$$

$$\text{con} \quad r = \min(m,n)$$

tales que

$$A = U \, \Sigma \, V^t.$$

Demostración.

Por uno de los Lemas anteriores sabemos que existen x_1, y_1 tales que

$$Ax_1 = \sigma_1 \, y_1$$

$$\text{con} \quad \sigma_1 = \|A\|$$

Por otro Lema, se sabe que existen U_1 de $m \times m$ y V_1 de $n \times n$, ortogonales, tales que

$$U_1^t \, y_1 = e_1^{(m)}$$

$$V_1^t\, x_1 = e_1^{(n)}$$

Al sustituir los valores

$$x_1 = V_1\, e_1^{(n)}$$

$$y_1 = U_1\, e_1^{(m)}$$

en

$$Ax_1 = \sigma_1\, y_1$$

Obtenemos

$$AV_1\, e_1^{(n)} = \sigma_1\, U_1\, e_1^{(m)}$$

es decir

$$(U_1^t\, AV_1)\, e_1^{(n)} = \sigma_1\, e_1^{(m)}$$

Esto implica que

$$A_2 = U_1^t\, AV_1$$

tiene la estructura siguiente:

$$A_2 = \begin{array}{cc} \begin{bmatrix} \sigma_1 & w^t \\ 0 & B_2 \end{bmatrix} & \begin{array}{l} 1 \\ m-1 \end{array} \\ \begin{array}{cc} \;\;1 & \;\;n-1 \end{array} & \end{array}$$

y como U_1 y V_1 son ortogonales, tenemos que

$$\sigma_1 = \|A\| = \|A_2\|$$

debido a que las matrices ortogonales conservan la norma de una matriz.

Continuando con la demostración, haremos ver que

$$w = 0$$

pues de otra forma no se obtendría la estructura diagonal de Σ y concluiremos que $\|B\| < \sigma$,

Supongamos pues que $w \neq 0$, entonces consideremos

$$z = A_2\ u$$

$$\text{con} \quad u = \begin{bmatrix} \sigma_1 \\ w \end{bmatrix}$$

Esto equivale a escribir

$$z = \begin{bmatrix} \sigma_1 & w^t \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + \|w\|^2 \\ B_2\ w \end{bmatrix}$$

y como

$$\|A_2\| = \max_{x \neq 0} \frac{\|A_2\ x\|}{\|x\|}$$

tenemos

$$\frac{\|z\|^2}{\|u\|^2} = \frac{\|A_2\ u\|^2}{\|u\|^2}$$

$$= \frac{(\sigma_1^2 + \|w\|^2)^2 + \|B_2\ w\|^2}{\sigma_1^2 + \|w\|^2}$$

$$= (\sigma_1^2 + \|w\|^2) + \frac{\|B_2\ w\|^2}{\sigma_1^2 + \|w\|^2}\ .$$

Como

$$\frac{\|z\|^2}{\|u\|^2} > \sigma_1^2 + \|w\|^2 > \sigma_1^2$$

$$\text{si} \quad \|w\| \neq 0$$

Tendríamos que

$$\frac{\|A_2\, u\|}{\|u\|} > \sigma_1 \qquad \text{si } \|w\| \neq 0$$

es decir $\quad \|A\| = \|A_2\| > \sigma_1 \quad$ si $\|w\| \neq 0$

lo cual contradice la definición de

$$\sigma_1 = \|A\|.$$

Por consiguiente $\quad w = 0.$

De donde

$$A_2 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B_2 \end{bmatrix}$$

Además, de lo anterior se obtiene directamente que

$$\|B_2\| < \sigma_1$$

Continuaremos ahora la demostración por inducción.

Supongamos que hemos obtenido

$$A_k = \begin{array}{cc} \left[\begin{array}{c:c} \begin{matrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{k-1} \end{matrix} & 0 \\ \hdashline 0 & B_k \end{array}\right] & \begin{array}{l} k-1 \\ \\ m-k+1 \end{array} \\ \begin{array}{cc} k-1 & \qquad n-k+1 \end{array} & \end{array}$$

donde

$$\sigma_1 > \sigma_2 > \ldots > \sigma_{k-1}$$

y

$$\| B_k \| < \sigma_{k-1}$$

y

$$A_k = U_{k-1} \; A_{k-1} \; V_{k-1}$$

Procediendo como en el caso inicial, existen $\tilde{U}_k$ y $\tilde{V}_k$ ortogonales tales que

$$\tilde{U}_k^t \; B_k \; \tilde{V}_k = \begin{array}{cc} \left[\begin{array}{c:c} \sigma_k & 0 \\ \hdashline 0 & B_{k+1} \end{array}\right] & \begin{array}{l} 1 \\ m-k \end{array} \\ \begin{array}{cc} 1 & \qquad n-k \end{array} & \end{array}$$

con

$$\sigma_k = \| B_k \| > \| B_{k+1} \|$$

Resta ahora obtener las U_k y V_k correspondientes.

Definiremos entonces,

$$V_k = \left[\begin{array}{c:c} I_{k-1} & 0 \\ \hdashline 0 & \tilde{V}_k \end{array}\right] ; \quad U_k = \left[\begin{array}{c:c} I_{k-1} & 0 \\ \hdashline 0 & \tilde{U}_k \end{array}\right]$$

y al aplicárselas a A_k, obtenemos

$$U_k^t \, A_k \, V_k = \left[\begin{array}{cccc:c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_k & \\ \hdashline & 0 & & & B_{k+1} \end{array}\right].$$

de donde podemos concluir que existen U y V ortogonales tales que

$$A = U \operatorname{diag}[\sigma_1, \sigma_2, \ldots, \sigma_r] \, V^t$$

con

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$$

para $r = \min(m,n)$. ∎

De la demostración anterior se deriva el siguiente:

Corolario.

La norma de A es igual al valor singular más grande de A.

$$\|A\| = \sigma_1 .$$

INTERPRETACION GEOMETRICA DE LOS VALORES SINGULARES.

Veremos ahora que los valores singulares son las longitudes de los diferentes ejes de un elipsoide de r dimensiones, donde

$$\sigma_1 = \sigma_{max}$$

$$\sigma_r = \sigma_{min}$$

son las longitudes del eje mayor y del eje menor respectivamente.

Para lograr lo anterior, haremos ver que la "bola unitaria" en $\mathbb{R}^n$ se transforma, bajo A, en un elipsoide r-dimensional, que es sólido si $r < n \leq m$.

Esto es, el conjunto

$$B = \{x \mid \|x\| = 1\}$$

que conocemos como la "bola unitaria" en $\mathbb{R}^n$, se transforma en

$$Z = \{z \mid z = Ax \text{ con } \|x\| = 1\}$$

Para ver que este conjunto es un elipsoide, emplearemos la Descomposición Singular de A:

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

entonces, $z \in Z$ puede escribirse como

$$z = U \Sigma V^t x$$

Ahora, si tomamos

$$y = V^t x$$

tenemos que por la ortogonalidad de V,

$$\|y\| = \|x\| = 1$$

y además

$$U^t z = \Sigma y$$

si consideramos ahora el conjunto

$$\tilde{Z} = \{u \mid u = U^t z, \ \ z \in Z\}$$

observamos que tiene las mismas propiedades geométricas que Z debido a que U es ortogonal. Así que usaremos por comodidad al conjunto $\tilde{Z}$.

Tenemos $\qquad u = \Sigma y \quad$ con $\|y\| = 1$

como

$$\Sigma = \begin{array}{cc} \left[\begin{array}{c:c} \tilde{\Sigma} & 0 \\ \hdashline 0 & 0 \end{array}\right] & \begin{array}{c} r \\ m-r \end{array} \\ \begin{array}{cc} r & \quad n-r \end{array} & \end{array}$$

con

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}$$

Si hacemos ahora

$$u = \begin{array}{c} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \end{array} \begin{array}{c} r \\ m-r \end{array}$$

y

$$y = \begin{array}{c} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \end{array} \begin{array}{c} r \\ n-r \end{array}$$

resulta que

$$u = \begin{bmatrix} \tilde{\Sigma}\, y_1 \\ 0 \end{bmatrix}$$

es decir

$$w_1 = \tilde{\Sigma}\, y_1$$

$$w_2 = 0$$

Para despejar a y_1, aprovechamos la estructura diagonal de $\tilde{\Sigma}$

$$y_1 = \tilde{\Sigma}^{-1}\, w_1$$

Y al aplicar normas

$$\|y_1\| = \|\tilde{\Sigma}^{-1} w_1\|$$

tenemos

$$\|y_1\|^2 \leq 1$$

ya que

$$\|y_1\|^2 + \|y_2\|^2 = 1$$

Entonces

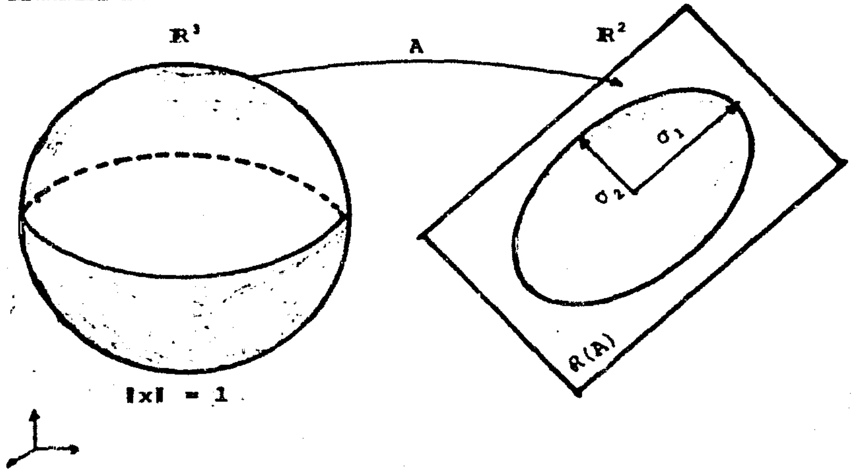
$$\sum_{i=1}^{r} \frac{w_i^2}{\sigma_i^2} \leq 1$$

Y se tiene que el conjunto $\tilde{Z}$ está dada por

$$\tilde{Z} = \left\{ u = \begin{array}{c} \begin{bmatrix} w_1 \\ 0 \end{bmatrix} \end{array} \begin{array}{l} r \\ m-r \end{array} \Big/ \sum_{i=1}^{r} \frac{w_i^2}{\sigma_i^2} \leq 1 \right\}$$

En donde se ve que las σ_i's con las longitudes de los ejes de un elipsoide de r dimensiones.

Gráficamente la situación es la siguiente con $n = 3$ y $r = 2$:



LA SEUDOINVERSA.

Hemos visto ya que el problema de "Encontrar la $x \in \mathbb{R}^n$ de norma mínima que minimice a

$$\|b - Ax\|^2 \text{ ''}.$$

tiene solución únic para cada $b \in \mathbb{R}^m$ con A una matriz fija de $m \times n$. Esto significa que a cada $b \in \mathbb{R}^n$ le podemos asociar la solución de norma mínima, x_{min}, del problema anterior.

Como x_{min} depende de la matriz A y del vector b, la podemos escribir provisionalmente como una función de dos parámetros: $x_{min}(A, b)$.

Ahora bien, si consideramos a x_{min} como una función de b con A fija tenemos

$$x_{min}(A, \cdot) : \mathbb{R}^m \to \mathbb{R}^n$$

Y así $x_{min}(A, b)$ es la solución del problema anterior. Notemos que $x_{min}(A, b)$ depende de manera muy simple de b, es decir $x_{min}(A, b)$ es una función lineal de b.

En efecto, en una sección anterior hicimos ver que

$$x_{min}(A, b) = B^t (B B^t)^{-1} (Q^t Q)^{-1} Q^t b$$

donde Q y B son tales que

$$A = Q B$$

y $\quad \text{rango } (A) = \text{rango } (Q) = \text{rango } (B) = r$

Esta expresión sirve para confirmarnos que $x_{min}(A, b)$ es una función lineal de b.

De aquí que podemos denotar

$$x_{min}(A, b) = A^+ b$$

en donde A^+ es una matriz de $n \times m$ que depende unívocamente de A, y está dada por

$$A^+ = B^t(B B^t)^{-1}(Q^t Q)^{-1}Q^t$$

Obsérvese que si $b \in \mathcal{R}(A)$, entonces

$$Ax_{min}(A,b) = A A^+ b = b$$

y en general

$$Ax_{min}(A,b) = A A^+ b = P_{\mathcal{R}(A)} b$$

de donde

$$A A^+ = P_{\mathcal{R}(A)}$$

Ya con esta discusión se puede presentar la siguiente.

<u>Definición.</u>

La seudoinversa A^+ de A es la función lineal, tal que

$$A^+ : \mathbb{R}^m \to \mathbb{R}^n$$

asocia a cada $b \in \mathbb{R}^m$, el vector

$$x_{min} = A^+ b$$

que tiene las siguientes propiedades:

i) $\min\limits_{x} \| b - Ax \|^2 = \| b - Ax_{min} \|^2$

ii) Si $\tilde{x} \neq x_{min}$ y $\tilde{x}$ satisface la propiedad anterior, entonces

$$\| x_{min} \|^2 < \| \tilde{x} \|^2$$

De la definición anterior se concluye el resultado siguiente:

Teorema.

Si
$$A = Q B$$
con
$$\text{rango } (A) = \text{rango } (Q) = \text{rango } (B) = r$$
entonces la seudoinversa de A está dada por

$$A^+ = B(B B^t)^{-1} (Q^t Q)^{-1} Q^t.$$

Con el desarrollo que presentamos a continuación, se obtiene una nueva caracterización de A^+, pero, esta vez, por medio de la Descomposición Singular.

Consideremos una vez más, el problema:

$$\min_{x} \| b - Ax \|^2$$

con $A_{m\times n}$ y rango $(A) = r < n$, entonces si

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

es la Descomposición Singular de A, tenemos que:

$$\begin{aligned}\min_x \|b - Ax\|^2 &= \min_x \|U(\Sigma V^t x - U^t b)\|^2 \\ &= \min_x \|\Sigma V^t x - U^t b\|^2 \\ &= \min_y \|\Sigma y - c\|^2\end{aligned}$$

en donde

$$V^t x = y$$

$$U^t b = c$$

ya que U y V son ortogonales y conservan la norma.

Como

$$y = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}; \quad c = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \text{ y } \Sigma = \begin{array}{cl} \begin{bmatrix} \sigma_1 & & & & & 0 \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & \cdots & \\ & & & \vdots & 0 & \\ 0 & & & & & \ddots \; 0 \end{bmatrix} & m-r \\ \qquad\qquad\qquad n-r & \end{array}$$

al calcular y_{min} de

$$\min_y \|\Sigma y - c\|^2$$

se obtiene

$$\| \Sigma y - c \|^2 = \sum_{k=1}^{r} (\sigma_k \ \eta_k - \gamma_k)^2 + \sum_{k=r+1}^{n} \gamma_k^2$$

en donde las η_k que minimizan esta expresión son aquéllas que satisfacen

$$\sigma_k \ \eta_k = \gamma_k$$

$$\text{para} \quad k = 1,2,\ldots,r$$

por lo que las $\tilde{y}$'s que minimizan, están dadas por

$$\tilde{y} = \begin{bmatrix} \frac{\gamma_1}{\sigma_1} \\ \frac{\gamma_2}{\sigma_2} \\ \vdots \\ \frac{\gamma_r}{\sigma_r} \\ \eta_{r+1} \\ \eta_{r+2} \\ \vdots \\ \eta_n \end{bmatrix}$$

con $\eta_k \in \mathbb{R}$ para $k = r+1,\ r+2,\ldots,n$ arbitrarias y entonces el vector y_{min} de norma mínima será aquél en el que

$$\eta_k = 0$$

$$\text{para} \quad k = r+1,\ r+2, \ldots, n$$

es decir

$$y_{min} = \begin{bmatrix} \frac{\gamma_1}{\sigma_1} \\ \frac{\gamma_2}{\sigma_2} \\ \vdots \\ \frac{\gamma_r}{\sigma_r} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

y la podemos escribir como

$$y_{min} = \Sigma^+ c$$

donde Σ^+ está dada por

$$\Sigma^+_{nxm} = \left[\begin{array}{cccc|c} \sigma_1^{-1} & & & & 0 \\ & \sigma_2^{-1} & & & \\ & & \ddots & & \\ & & & \sigma_r^{-1} & \\ \hline & & & & \\ & 0 & & & 0 \end{array}\right]$$

Pero como

$$V^t x = y$$

y

$$U^t\, b = c$$

se tiene que

$$V^t\, x_{min} = \Sigma_c^+ = y_{min}$$

de donde

$$V^t\, x_{min} = \Sigma^+\, U^t\, b$$

y entonces

$$x_{min} = V\, \Sigma^+\, U^t\, b.$$

Como

$$x_{min} = A^+\, b$$

obtenemos

$$A^+ = V\, \Sigma^+\, U^t$$

que es la caracterización de la Seudoinversa de A, en términos de la Descomposición Singular de A.

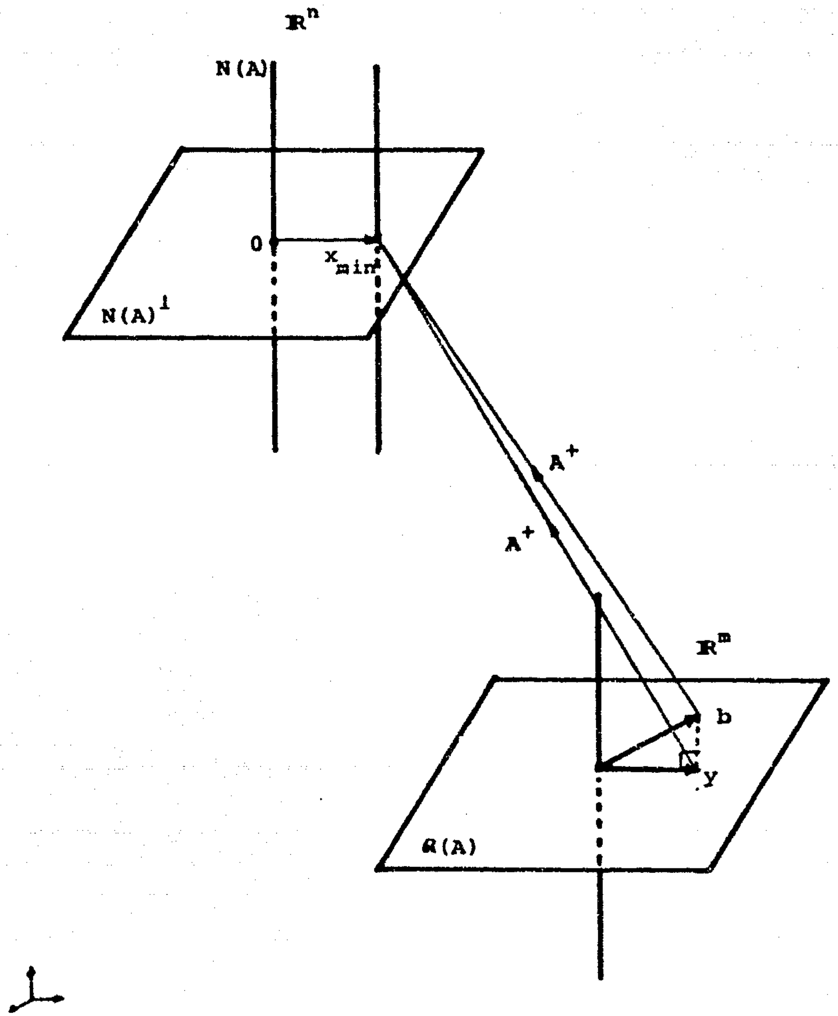
Para comprender el concepto de Seudoinversa, $A^+ : \mathbb{R}^m \to \mathbb{R}^n$, nos auxiliaremos de la figura siguiente que nos muestra el efecto geométrico de A^+.

En primer lugar, proyectamos a un vector b en $\mathbb{R}^m$ sobre el $\mathcal{R}(A)$, obtenido

$$y = P_{\mathcal{R}(A)}\, b$$

y después encontramos al único vector x_{min} en $N(A)^{\perp} = \mathcal{R}(A^t)$ que resuelve al sistema

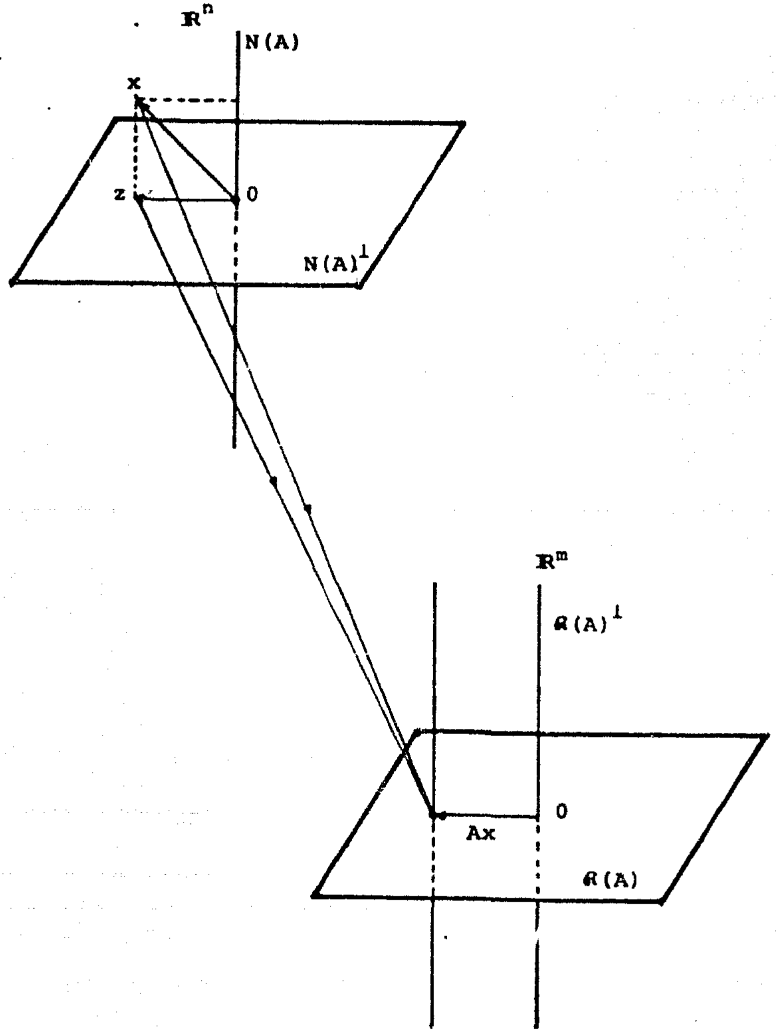
$$Ax_{min} = y$$



Y, en forma análoga, la siguiente figura ilustra el efecto geométrico de la matriz A, $A: \mathbb{R}^n \to \mathbb{R}^m$. Proyectamos, entonces, a un vector x sobre $N(A)^{\perp} = \mathcal{R}(A^t)$ para obtener

$$z = P_{N(A)^{\perp}}\, x$$

Al aplicar A a z y x resulta $Az = Ax$



PROPIEDADES BASICAS DE LA SEUDOINVERSA.

1) Si $A:\mathbb{R}^n \to \mathbb{R}^m$, entonces

i) $A A^+ = P_{\mathcal{R}(A)}$

ii) $A^+ A = P_{\mathcal{R}(A^t)} = P_{N(A)^\perp}$

iii) $I - A A^+ = P_{\mathcal{R}(A)^\perp}$

iv) $I - A^+ A = P_{N(A)}$

Prueba.

i) $A \quad A^+ = P_{\mathcal{R}(A)}$

Sea $z \in \mathbb{R}^m$. Como $A^+ z$ es la solución de norma mínima de

$$\min_x \| z - Ax \|^2$$

y también la solución de

$$\min_x \| z_1 - Ax \|^2$$

donde

$$z_1 = P_{\mathcal{R}(A)} z.$$

pero

$$\min_x \| z_1 - Ax \|^2 = \| z_1 - A(A^+ z) \|^2 = 0$$

porque

$$z_1 \in \mathcal{R}(A).$$

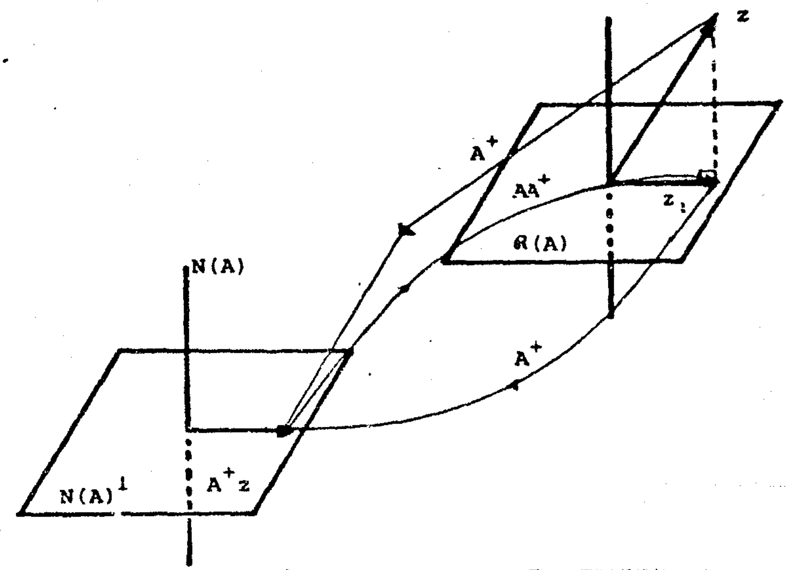
Entonces

$$\begin{aligned} A A^+ z &= z_1 \\ &= P_{\mathcal{R}(A)} z \end{aligned}$$

para toda $z \in \mathbb{R}^m$

de donde $$A\,A^{+} = P_{\mathcal{R}(A)}. \quad \blacksquare$$

Para una mejor comprensión de esta prueba consideremos la figura siguiente:



Las pruebas para (ii), (iii) y (iv) se derivan de la anterior.

2) Si $A : \mathbb{R}^n \to \mathbb{R}^m$, entonces

i) $A = A\,A^{+}\,A$

ii) $A^{+} = A^{+}\,A\,A^{+}$

<u>Prueba.</u>

i) $A = A\,A^{+}\,A$

Se tiene, por el resultado anterior que

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

y si lo anterior, se aplica a A, tenemos

$$A\,A^{+}\,A = P_{\mathcal{R}(A)}A = A$$

de donde

$$A\,A^{+}\,A = A.$$

∎

ii) $A^{+} = A^{+}\,A\,A^{+}$

Por el resultado anterior

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

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

de donde

$$A^{+}\,A\,A^{+} = A^{+}$$

Para terminar con esta sección presentaremos lo concerniente a la norma de la seudoinversa:

$$\|A^{+}\| = \max_{\|x\|=1} \|A^{+}x\| = \frac{1}{\sigma_r}$$

ya que si

$$x \in \mathbb{R}^{m}$$

$$\|A^{+}x\|^{2} = \|V\,\Sigma^{+}\,U^{t}x\|^{2}$$

$$= \|\Sigma^{+}y\|^{2}$$

con

$$U^{t}\,x = y$$

es decir

$$\| A^{+} x \|^{2} = (\frac{\eta_1}{\sigma_1})^{2} + (\frac{\eta_2}{\sigma_2})^{2} + \ldots + (\frac{\eta_r}{\sigma_r})^{2}$$

como

$$\| A^{+} x \|^{2} \leq (\frac{1}{\sigma_r})^{2} \| y \|^{2} = (\frac{1}{\sigma_r})^{2} \| x \|^{2}$$

ya que

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$$

y si hacemos

$$y_0 = \begin{array}{cc} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} & \begin{array}{c} \\ \\ \\ \longleftarrow r \\ \\ \\ \\ \\ \\ \end{array} \end{array}$$

y

$$y_0 = U^{t} x_0$$

tenemos

$$x_0 = U y_0$$

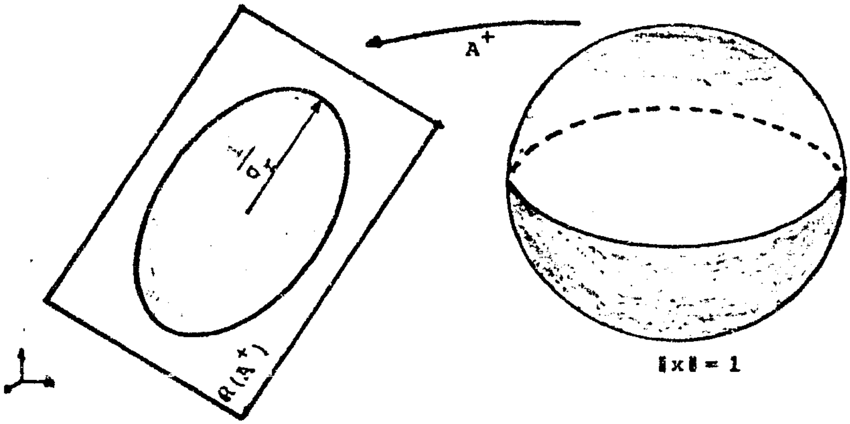
Entonces

$$\| A^{+} x_0 \|^{2} = (\frac{1}{\sigma_r})^{2} \| x_0 \|^{2}$$

de donde

$$\| A^{+} \| = \frac{1}{\sigma_r}$$

Veamos gráficamente en $\mathbb{R}^3$ el efecto de A^{+} en la bola unitaria:



En este capítulo hemos presentado tres conceptos:

La Descomposición Singular

La Norma de una matriz

y La Seudoinversa.

Cada uno de ellos será de utilidad durante el desarrollo de este trabajo.

Cabe señalar, que el concepto de norma de una matriz servirá, en su momento, para cuantificar el efecto que cambios en los elementos de la matriz A y el vector B pueden producir en la solución de norma mínima del problema de Cuadrados Mínimos:

$$\min_x \|b - Ax\|^2 .$$

C A P I T U L O   IV

TEORÍA DE PERTURBACIÓN

INTRODUCCIÓN.

En esta sección estudiaremos cómo afectan las alteraciones en los datos a la solución de norma mínima del problema

$$\min_{x} \|Ax - b\|^2 .$$

es decir estudiaremos la sensibilidad de la solución.

El conocimiento de dicha sensibilidad es muy importante desde el punto de vista práctico debido a que nos indica la manera cómo los errores en las observaciones y los errores de redondeo durante el proceso de cálculo afectan a la solución.

En este trabajo no establecemos diferencias entre los errores y sólo los interpretamos como alteraciones en los datos, es decir, en la matriz A y el vector b.

Nos interesa, además, establecer las condiciones bajo las cuales la solución del problema alterado es cercano a la del problema original. Es decir, queremos saber bajo qué condiciones la solución del problema de Cuadrados Mínimos varía continuamente con respecto a los datos; para que, cuando esto suceda podamos estimar los cambios en la solución en función de los cambios en los datos.

Podemos diferenciar tres tipos de alteraciones en los datos del problema de Cuadrados Mínimos:

i) Cuando cambios o alteraciones modifican sólamente al vector b, y en la práctica lo que realmente tenemos es un vector $\tilde{b}$.

ii) Cuando sólo A es alterada y tenemos $\tilde{A}$.

iii) Cuando los dos casos anteriores ocurren al mismo tiempo.

De acuerdo con esta clasificación de las alteraciones podemos dividir en tres casos o preguntas nuestro estudio:

1) Si $\tilde{b}$ difiere poco de b, las soluciones de norma mínima de

$$\min_x \|Ax - b\|^2 \quad y \quad \min_x \|Ax - \tilde{b}\|^2$$

¿diferirán poco también?

2) ¿Será válido suponer que si los elementos de $\tilde{A}$ y A son ligeramente diferentes, las soluciones de norma mínima de los problemas

$$\min_x \|Ax - b\|^2 \quad y \quad \min_x \|\tilde{A}x - b\|^2$$

serán también un poco diferentes?

3) ¿Cómo afectará a la solución de

$$\min_x \|Ax - b\|^2$$

cambios en los elementos de A y de b?

Antes de responder a estas preguntas, introduciremos algunos conceptos que después nos serán útiles.

Si

$$\tilde{A} = A + E$$

y $\|E\|$ es pequeña con respecto a $\|A\|$, decimos que $\tilde{A}$ es la matriz A perturbada por E.

Similarmente, si

$$\tilde{b} = b + e$$

y $\|e\|$ es pequeña con respecto a $\|b\|$, decimos que $\tilde{b}$ es el vector b perturbado por e.

Según lo anterior, la Teoría de la Perturbación es aquélla que nos explica la relación que puede existir entre la solución de un problema y la solución de ese mismo problema pero perturbado.

Durante los desarrollos que iremos presentando para responder a las preguntas anteriores, pretendemos motivar algunos resultados básicos sobre Teoría de la Perturbación para el problema de Cuadrados Mínimos.

Utilizaremos, entre otros conceptos, al de Seudoinversa, pues se trata de un concepto teórico que proporciona una expresión compacta para la solución del problema de Cuadrados Mínimos en general. Es decir, con ella, obtenemos una expresión para la solución, independientemente del hecho de que la matriz tenga o no rango completo. Emplearemos también el concepto de norma para interpretar el significado de "cercanía" entre matrices y entre vectores.

Comenzaremos con el caso cuando $\tilde{b}$ difiere poco de b; la respuesta a la pregunta planteada la proporciona el siguiente:

Teorema.

Sea

$$x = A^+ b$$

la solución de norma mínima del problema.

$$\min_z \|Az - b\|^2$$

y sea

$$\tilde{x} = A^+ \tilde{b}$$

la solución de norma mínima del problema

$$\min_z \|Az - \tilde{b}\|^2$$

Entonces

$$\tilde{x} \to x$$

cuando

$$\tilde{b} \to b$$

Es decir, si $\tilde{b}$ es cercana a b, entonces $\tilde{x}$ es cercana a x.

Demostración.

Calculemos

$$\|\tilde{x} - x\|$$

para ver qué tan cerca se encuentra una solución de la otra:

Así, entonces

$$\|\tilde{x} - x\|^2 = \|A^+ \tilde{b} - A^+ b\|^2$$

$$= \|A^+ (\tilde{b} - b)\|^2$$

En donde se puede ver que cuando

$$\tilde{b} \to b$$

$$\|\tilde{x} - x\|^2 \to 0$$

■

Esto quiere decir que los errores pequeños en b no causan problemas en la solución de norma mínima.

Pasamos, ahora, a considerar el caso en el que la matriz A es perturbada:

Sean

$$x = A^+ b$$

$$\tilde{x} = \tilde{A}^+ b$$

donde

$$\tilde{A} = A + E,$$

entonces, si queremos ver qué tan cerca está $\tilde{x}$ de x, calculamos:

$$\tilde{x} - x = \tilde{A}^+ b - A^+ b$$

$$= (\tilde{A}^+ - A^+) b$$

y, al aplicar normas, obtenemos:

$$\|\tilde{x} - x\| \leq \|\tilde{A}^{+} - A^{+}\| \cdot \|b\|$$

esta desigualdad nos plantea una cuestión que resulta interesante:

¿Si $\|E\| \to 0$

entonces

$\|\tilde{A}^{+} - A^{+}\| \to 0$ también?

Y... ¿qué quiere decir ésto?

Pues que si ocurriera lo anterior sería fácil para nosotros saber la "cercanía" entre $\tilde{x}$ y x.

Sin embargo, en este momento, no tenemos información suficiente para responder a esa pregunta.

Para aclarar esta cuestión presentaremos a continuación dos ejemplos simples, con los que estudiaremos los efectos que las perturbaciones producen en A^{+} y x, al resolver el problema de

$$\min \|\tilde{A}x - b\|^2$$

Primer Ejemplo.

Sea $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ y consideremos a A_1 una matriz muy cercana a A, dada por

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}$$

en donde $\varepsilon \in \mathbb{R}$ y $|\varepsilon| < 1$

Sea $$b = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Para resolver el problema

$$\min_x \|Ax - b\|^2$$

proyectamos a b sobre $\mathcal{R}(A)$ que coincide con el eje de las ab_ bisas en $\mathbb{R}^2$

y como $$x = A^+ b$$

tenemos que $$x = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

y $$A^+ = A$$

En cuanto a A_1, tenemos lo siguiente

$$A_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\varepsilon} \end{bmatrix}$$

$$= A_1^{-1}$$

de donde $$x_1 = A_1^+ b$$

$$= \begin{bmatrix} \beta_1 \\ \dfrac{\beta_2}{\varepsilon} \end{bmatrix}$$

y la diferencia es

$$\| x_1 - x \|^2 = \left(\frac{\beta_2}{\varepsilon}\right)^2$$

en donde se puede ver que cuando $\varepsilon \to 0$,

$$\| x_1 - x \|^2 \to \infty$$

Lo que quiere decir que x_1 se aleja indefinidamente de x cuando $\varepsilon \to 0$.

En forma análoga se obtiene

$$\| A_1^+ - A^+ \|^2 = \frac{1}{\varepsilon^2}$$

en donde, otra vez, se puede ver que cuando $\varepsilon \to 0$, A_1^+ se aleja indefinidamente de A^+, es decir,

$$\| A_1^+ - A^+ \|^2 \to \infty$$

cuando $\varepsilon \to 0$

y, de lo anterior, se ve que

$$x_1 \neq x$$

cuando $$\varepsilon \to 0$$

$$A_1^+ \neq A^+$$

cuando $\qquad \varepsilon \to 0$

Pero ... ¿a qué se debe esto?

Como se verá más adelante, la causa de que x_1 no esté "cerca" de x y A_1^+ no esté "cerca" de A^+ se encuentra en el hecho de que la perturbación altera la dimensión de $\mathcal{R}(A)$, ya que

$$2 = \dim(\mathcal{R}(A_1)) \neq \dim(\mathcal{R}(A)) = 1$$

para toda $\varepsilon \neq 0$

Segundo Ejemplo.

Sea A, definida en el ejemplo anterior y

$$A_2 = \begin{bmatrix} 1 & \varepsilon \\ 0 & 0 \end{bmatrix}$$

Claramente, A_2 es cercano a A, entonces

$$A_2 x = \begin{bmatrix} \xi_1 + \varepsilon\xi_2 \\ 0 \end{bmatrix}$$

donde

$$x = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

y tenemos que

$$\min_{x} \| A_2 x - b \|^2 = \min_{x} [(\xi_1 + \varepsilon\xi_2) - \beta_1]^2 + \beta_2^2$$

pero como

$$\beta_1 = \xi_1 + \varepsilon\xi_2$$

al tomar $\qquad \alpha = \xi_2$

resulta $\qquad \xi_1 = \beta_1 - \varepsilon\alpha$

De donde al minimizar

$$f(\alpha) = (\beta_1 - \varepsilon\alpha)^2 + \alpha^2$$

tenemos

$$\alpha = \frac{\beta_1 \ \varepsilon}{1 + \varepsilon^2} = \xi_2$$

de donde

$$\xi_1 = \beta_1 (1 - \frac{\varepsilon^2}{1+\varepsilon^2})$$

$$= \beta_1 (\frac{1}{1+\varepsilon^2})$$

de aquí que

$$x_2 = A_2^+ b$$

$$= \begin{bmatrix} \beta_1 \ (\frac{1}{1+\varepsilon^2}) \\ \\ \beta_1 \ (\frac{\varepsilon}{1+\varepsilon^2}) \end{bmatrix}$$

es decir
$$x_2 = \begin{bmatrix} \frac{1}{1+\varepsilon^2} & 0 \\ \\ \frac{\varepsilon}{1+\varepsilon^2} & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \\ \beta_2 \end{bmatrix}$$

de donde
$$A_2^+ = \begin{bmatrix} \frac{1}{1+\varepsilon^2} & 0 \\ \\ \frac{\varepsilon}{1+\varepsilon^2} & 0 \end{bmatrix}$$

y la diferencia entre x_2 y x es:

$$\| x_2 - x \|^2 = \frac{\beta_1^2 \ \varepsilon^2 \ (\varepsilon^2 + 1)}{(1+\varepsilon^2)^2}$$

$$= \frac{\beta_1^2 \ \varepsilon^2}{(1+\varepsilon^2)}$$

de donde vemos que cuando $\varepsilon \to 0$

$$\| x^2 - x \|^2 \to 0$$

es decir,

x_2 se asemeja a x cuando $\varepsilon \to 0$.

En cuanto al error en A_2^+ se tiene lo siguiente:

$$\|A_2^+ - A^+\|^2 = \sup_x \frac{\|(A_2^+ - A^+)\ x\|^2}{\|x\|^2}$$

$$= \sup_{\xi_1,\xi_2} \frac{\left(\frac{-\varepsilon^2\xi_1}{1+\varepsilon^2}\right)^2 + \left(\frac{\varepsilon\xi_1}{1+\varepsilon^2}\right)^2}{\xi_1^2 + \xi_2^2}$$

$$= \frac{\varepsilon^2}{1+\varepsilon^2}$$

De donde

$$\|A_2^+ - A^+\|^2 \to 0$$

cuando $\varepsilon \to 0$. Lo que significa que cuando $\varepsilon \to 0$, A_2^+ se aseme ja a A^+.

Y así

$$x_2 = x$$

cuando $\varepsilon \to 0$

$$A_2^+ = A^+$$

cuando $\varepsilon \to 0$

Observemos que la perturbación en A, en este caso tiene la propiedad de que

$$1 = \dim\ (\mathcal{R}(A)) = \dim\ (\mathcal{R}(A_2))$$

De los ejemplos anteriores vale la pena remarcar lo siguiente:

Para que x_1 y x_2 estén "cerca" de x, y para que A_1 y A_2 difieran poco de A, es importante que la perturbación no afec te la dimensión de $\mathcal{R}(A)$.

Cuando una perturbación, por pequeña que sea, altera $\dim(\mathcal{R}(A))$ su efecto en la seudoinversa de la matriz perturbada es considerable.

Observamos también que cuando $\dim(\mathcal{R}(A))$ se conserva, como en el segundo ejemplo, la solución x_2 resulta ser cualitativamente la misma: x. Esto no ocurrió en el primer ejemplo cuando se alteró $\dim(\mathcal{R}(A))$ por la perturbación.

En base a esto diremos que una perturbación conserva el rango si

$$\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A_i)) \qquad i=1,2.$$

<u>Nota.</u>

La afirmación anterior tiene que ver con el hecho de que para que dos subespacios estén "cercanos" es necesario que ambos tengan la misma dimensión, como más adelante veremos:

De acuerdo con lo anterior concluiremos con lo siguiente:

Sólo se permitirán aquellas perturbaciones en A que no afecten $\dim(\mathcal{R}(A))$; esto con el objeto de que la seudoinversa $\tilde{A}^+$ de $\tilde{A} = A + E$ no esté muy alejada de A^+.

Para terminar con los casos que se plantearon, falta únicamente referirnos al caso cuando $\tilde{A}$ difiere poco de A y $\tilde{b}$

está muy cerca de b.

Sin embargo, es fácil darse cuenta que este caso se reduce al anterior ya que las alteraciones en b no afectan a la solución de norma mínima.

PERTURBACION PARA INVERSAS.

Como la inversa es un caso particular de la seudoinversa, estudiaremos este caso simple y estableceremos los resultados de tal forma que nos permitan su extensión al caso de seudoinversas en general.

Veamos lo que ocurre cuando la matriz identidad I es perturbada.

Tendremos entonces matrices de la forma

$$I - E$$

Después examinaremos el caso cuando las matrices son de la forma

$$A + E$$

Los resultados para el segundo caso se derivan del primero.

El siguiente es un resultado clásico que da una condición suficiente para que la matriz perturbada

$$I - E$$

donde E es una matriz de perturbación, sea no singular:

Lema de Banach.

Sea E una matriz de orden n y tal que $\|E\| < 1$.

Entonces

$$I_n - E$$

es no singular y

$$\|(I_n - E)^{-1}\| < \frac{1}{1-\|E\|} .$$

Demostración.

Sea $x \in \mathbb{R}^n$ y $x \neq 0$, entonces

$$\|(I_n - E)x\| = \|x - Ex\|$$

$$\geq \|x\| - \|Ex\|$$

$$\geq \|x\| - \|E\| \cdot \|x\|$$

$$\geq (1 - \|E\|)\|x\| > 0$$

ya que

$$(1 - \|E\|) > 0$$

y como $x \neq 0$ tenemos que

$$(I - E)x \neq 0$$

de donde

$$(I - E) \text{ es no singular.}$$

Ahora, de la ecuación

$$(I-E)(I-E)^{-1} = I$$

se tiene que

$$(I-E)^{-1} = I + E(I-E)^{-1}$$

entonces

$$\|(I-E)^{-1}\| < \|I\| + \|E\| \cdot \|(I-E)^{-1}\|$$

como

$$\|I\| = 1$$

se tiene

$$\|(I-E)^{-1}\| < \frac{1}{1-\|E\|} \quad . \qquad \blacksquare$$

Al interpretar en términos de valores singulares el Lema de Banach, tenemos:

<u>Lema.</u>

Si E es una matriz de orden n y ε_1 su valor singular más grande tal que

$$\|E\| = \varepsilon_1 < 1$$

entonces

$$(I-E) \text{ es no singular, y}$$

$$0 < (1-\varepsilon_1) < \lambda_n$$

donde λ_n es el valor singular más pequeño de $(I-E)$.

Un corolario del Lema de Banach establece una condición suficiente para que

$$\tilde{A} = A + E$$

sea no singular donde E es una matriz de perturbación.

Corolario.

Sean A_{nxn} una matriz no singular y E_{nxn} otra matriz, tales que

$$\varepsilon_1 < \sigma_n$$

donde ε_1 es el valor singular más grande de E y σ_n el valor singular más pequeño de A.

Entonces,

$$A + E \text{ es no singular, y}$$

$$\tilde{\sigma}_n \geq \sigma_n - \varepsilon_1 > 0$$

donde $\tilde{\sigma}_n$ es el valor singular más pequeño de $(A + E)$.

Demostración.

Sea

$$(A + E) = A(I + A^{-1} E)$$

si tomamos

$$B = - A^{-1} E$$

entonces

$$\beta_1 = \|B\| \leq \|A^{-1}\| \cdot \|E\| = \frac{\varepsilon_1}{\sigma_n} < 1$$

y

$$A + E = A(I - B)$$

Por el Lema de Banach $(I - B)$ es no singular, por lo tanto

$$(A+E) \text{ es no singular}$$

y

$$(A+E)^{-1} = (I-B)^{-1} A^{-1}$$

Por consiguiente

$$\|(A+E)^{-1}\| = \|(I-B)^{-1} A^{-1}\|$$

$$< \frac{\frac{1}{\sigma_n}}{1 - \frac{\varepsilon_1}{\sigma_n}}$$

de donde

$$\frac{1}{\tilde{\sigma}_n} < \frac{1}{\sigma_n - \varepsilon_1}$$

es decir

$$\tilde{\sigma}_n > \sigma_n - \varepsilon_1 > 0 .$$

∎

El resultado anterior se puede utilizar para hacer ver que dada una matriz A cuyos elementos son funciones continuas de una variable real ε, si A es invertible para $\varepsilon = \varepsilon_0$, entonces A es invertible para toda ε en una vecindad de ε_0.

El corolario siguiente establece lo anterior para el caso especial en que $\tilde{A}$ es de la forma

$$\tilde{A} = A + \varepsilon E$$

con $\|E\| = 1$.

<u>Corolario.</u>

Si A es no singular y

$$\tilde{A} = A + \varepsilon E$$

con $\|E\| = 1$ y $|\varepsilon| < \sigma_n$, entonces $\tilde{A}^{-1}$ es continua para toda ε en el intervalo

$$(-\sigma_n, \sigma_n).$$

Ahora, por medio del resultado siguiente, mostramos la importancia del valor singular más pequeño: σ_n.

Lema.

Dada $A_{n\times n}$ con rango $(A) = n$, existe $\tilde{A}$ matriz singular, tal que

$$\|A - \tilde{A}\| = \sigma_n.$$

Demostración.

Sea
$$A = U \Sigma V^t$$
la descomposición singular de A.

Entonces si definimos $\tilde{A}$ como

$$\tilde{A} = U \begin{bmatrix} \sigma_1 & & & & & 0 \\ & \sigma_2 & & & & \\ & & \sigma_3 & & & \\ & & & \ddots & & \\ & & & & \sigma_{n-1} & \\ 0 & & & & & 0 \end{bmatrix} V^t$$

se tiene que

$$\|A - \tilde{A}\| = \sigma_n .$$

El Lema siguiente es una reformulación del Lema de Banach.

Lema.

Sea A como en el Lema anterior, si $\tilde{A}$ es tal que

$$\|A - \tilde{A}\| < \sigma_n$$

entonces

$$\tilde{A} \text{ es invertible.}$$

Demostración.

$\tilde{A}$ es tal que

$$\tilde{A} = A + E$$

de donde

$$\|A - \tilde{A}\| = \|E\| = \varepsilon_1 < \sigma_n$$

por hipótesis.

Y como A es no singular, tenemos por un corolario anterior que

$$\tilde{A} \text{ es no singular.}$$

■

Corolario.

Sea A_{nxn} y no singular, entonces si B es singular de orden n, se tiene

$$\|A - B\| > \sigma_n$$

y

$$\sigma_n = \min\{\|A - B\| \mid B \text{ es singular}\}.$$

Los resultados anteriores equivalen a decir que la perturbación máxima que E puede producir en A, sin hacer que

$$\tilde{A} = A + E$$

sea no singular es aquélla tal que

$$\|E\| < \sigma_n .$$

Aquí terminamos con los resultados sobre perturbación en la matriz cuadrada A, en términos de los valores singulares de A.

Pasamos ahora al caso general, en donde se hará algo análogo para matrices A de m x n.

PERTURBACION DE VALORES CARACTERISTICOS.

Dedicaremos ahora nuestra atención al estudio de resultados sobre Perturbación en matrices de m x n.

En la sección anterior hemos visto cómo estos resultados se pueden establecer en términos de los valores singulares y en esta sección deduciremos resultados generales análogos.

Ahora bien, como los valores singulares de una matriz A, corresponden a las raíces cuadradas positivas de los valores característicos de $A^t A$, los resultados sobre perturbación en términos de valores singulares se basan en los correspondientes sobre valores característicos.

Por lo anterior, es conveniente que revisemos algunos resultados sobre perturbaciones en los valores característicos de matrices simétricas, las cuales servirán como antecedente para establecer los correspondientes sobre perturbación en los valores singulares.

Comenzaremos haciendo ver que la estructura de una matriz simétrica se refleja en el buen comportamiento de sus valores y vectores característicos al ser éstos perturbados. Esta cualidad de las matrices simétricas se manifiesta a través del Teorema Minimax. Este Teorema está relacionado con el cociente

$$\frac{y^t Ay}{y^t y}$$

cuyo rango de valores lo indica el resultado siguiente:

<u>Teorema.</u>

Si A es una matriz simétrica de orden n con valores característicos.

$$\lambda_1, \lambda_2, \ldots, \lambda_n$$

tales que

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n$$

y sus correspondientes vectores característicos,

$$x_1, x_2, \ldots, x_n$$

son ortonormales. Además, si

$$y \neq 0$$

entonces

$$\lambda_n < \frac{y^t\ Ay}{y^t\ y} < \lambda_1$$

para toda $y \in \mathbb{R}^n$.

<u>Demostración.</u>

Sea

$$y = \sum_{i=1}^{n} \rho_i\ x_i$$

con $\rho_i \in \mathbb{R}$

por ser $x_1, x_2, \ldots, x_n$ ortonormales

y

$$Ay = \sum_{i=1}^{n} \rho_i\ \lambda_i\ x_i$$

entonces

$$\frac{y^t\ Ay}{y^t\ y} = \frac{\sum_{i=1}^{n} |\rho_i|^2\ \lambda_i}{\sum_{i=1}^{n} |\rho_i|^2}$$

con

$$\sum_{i=1}^{n} |\rho_i|^2 = \|y\|^2 \neq 0$$

Al restar de esta expresión λ_n, tenemos

$$\frac{y^t\ Ay}{y^t\ y} - \lambda_n = \frac{\sum_{i=1}^{n-1} (\lambda_i - \lambda_n) |\rho_i|^2}{\sum_{i=1}^{n} |\rho_i|^2}$$

y como $(\lambda_i - \lambda_n) > 0$

para $i = 1,2,\ldots,n-1$

tenemos que

$$\frac{y^t\, Ay}{y^t\, y} > \lambda_n$$

Procediendo en forma análoga para λ_1 tenemos que

$$\lambda_n < \frac{y^t\, Ay}{y^t\, y} < \lambda_1 \qquad \blacksquare$$

Con este resultado estamos ya en condiciones de obtener una expresión "minimax" para cualquier valor característico λ_i.

Trataremos ahora de explicar lo que se entiende por expresión "minimax". Con el resultado anterior, tenemos ya expresiones para el mínimo y para el máximo de los valores característicos como lo establece el siguiente:

<u>Corolario.</u>

$$\lambda_1 = \max_{y\neq 0} \frac{y^t\, Ay}{y^t\, y}$$

$$\lambda_n = \min_{y\neq 0} \frac{y^t\, Ay}{y^t\, y}$$

Ahora, faltan expresiones para los demás valores característicos: λ_k

Estas se obtienen de la manera siguiente:

1)
$$\lambda_k = \max_{\substack{y \in S_k \\ y \neq 0}} \frac{y^t\, Ay}{y^t\, y}$$

donde S_k es el subespacio generado por los vectores característicos de A:

$$\{x_k,\ x_{k+1}, \ldots, x_n\}$$

Ahora bien, si S es un subespacio de la misma dimensión que S_k, entonces veremos que

$$\lambda_k < \max_{y \in S} \frac{y^t\, Ay}{y^t\, y}$$

de donde

$$\lambda_k = \min_{\dim(S) = n-k+1} \ \max_{\substack{y \in S \\ y \neq 0}} \frac{y^t\, Ay}{y^t\, y}$$

2)
$$\lambda_k = \min_{\substack{y \in \tilde{S}_k \\ y \neq 0}} \frac{y^t\, Ay}{y^t\, y}$$

donde $\tilde{S}_k$ es el subespacio generado por los vectores característicos de A

$$\{x_1, x_2, \ldots, x_k\}$$

Y si $\tilde{S}$ es un subespacio de la misma dimensión que $\tilde{S}_k$, entonces

$$\lambda_k > \min_{y \in \tilde{S}} \frac{y^t\, Ay}{y^t\, y}$$

de donde

$$\lambda_k = \max_{\dim(\tilde{S}) = k} \ \min_{\substack{y \in \tilde{S} \\ y \neq 0}} \frac{y^t\, Ay}{y^t\, y}$$

Estas expresiones minimax las deduce el Teorema Minimax Para enunciarlo introduciremos algunos conceptos que se utilizan en su presentación.

Definición.

Sea F_k la familia de los subespacios vectoriales de $\mathbb{R}^n$ de dimensión k. Dada una matriz Anxn simétrica, definimos a la función H_A

$$H_A : F_k \to \mathbb{R}$$

que está dada por

$$H_A(S) = \max_{\substack{x \in S \\ x \neq 0}} \frac{x^t Ax}{x^t x}$$

donde $S \in F_k$

Y definimos también la función h_A

$$h_A : F_k \to \mathbb{R}$$

dada por

$$h_A(S) = \min_{\substack{x \in S \\ x \neq 0}} \frac{x^t Ax}{x^t x}$$

donde $S \in F_k$.

Con estos elementos pasamos a presentar:

Teorema Minimax.

Sea A simétrica con valores característicos.

$$\lambda_1, \lambda_2, \ldots, \lambda_n$$

ordenados en la forma siguiente

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

Entonces

$$\lambda_k = \min_{S \in F_{n-k+1}} H_A(S) = \max_{S \in F_k} h_A(S)$$

Demostración.

Se probará primero que

$$\lambda_k = \min_{S \in F_{n-k+1}} H_A(S)$$

Usaremos primero un subespacio auxiliar

$$\tilde{S} = [x_1, x_2, \ldots, x_k]$$

generado por los primeros k vectores característicos de A.

Sea

$$x \in \tilde{S} \quad y \quad x \neq 0$$

entonces

$$\lambda_k \leq \frac{x^t Ax}{x^t x} \leq \lambda_1$$

Tomaremos ahora un subespacio de dimensión n-k+1 que tenga al menos un elemento diferente de cero en común con el anterior, esto caracterizará a λ_k como una cota inferior del máximo sobre este subespacio.

Así pues, sea

$$S \in F_{n-k+1}$$

entonces

$$S \cap \tilde{S} \neq \{0\}$$

ya que

$$\dim(S) + \dim(\tilde{S}) = n + 1$$

es decir, existe un vector

$$z \in S \cap \tilde{S}, \quad z \neq 0$$

Entonces

$$\lambda_k < \frac{z^t\ Az}{z^t\ z} < \lambda_1$$

De donde

$$\max_{x \in S} \frac{x^t\ Ax}{x^t\ x} = H_A(S) > \lambda_k \qquad \text{...(A)}$$

Continuamos ahora definiendo otro subespacio de dimensión n-k+1, con el cual obtendremos una expresión que señala a λ_k como el máximo sobre este tercer subespacio.

Sea entonces

$$S' = [x_k, x_{k+1}, \ldots, x_n]$$

si

$$x \in S', \ x \neq 0$$

tenemos que

$$\lambda_n < \frac{x^t\ Ax}{x^t\ x} < \lambda_k$$

entonces

$$\max_{x \in S'} \frac{x^t\ Ax}{x^t\ x} = H_A(S') < \lambda_k \qquad \text{...(B)}$$

Con las dos desigualdades anteriores (A) y (B) ya podemos concluir que λ_k es el mínimo de los máximos sobre los subespacios de dimensión $n-k+1$.

Es decir,

$$\lambda_k = \min_{S \in F_{n-k+1}} H_A(S)$$

Procederemos en forma análoga para probar que

$$\lambda_k = \max_{S \in F_k} h_A(S)$$

Sea entonces

$$\tilde{S} = [x_k, x_{k+1}, \ldots, x_n]$$

y sea $x \in \tilde{S}$, $x \neq 0$

entonces
$$\lambda_n < \frac{x^t\, Ax}{x^t\, x} < \lambda_k$$

Sea $S \in F_k$, entonces

$$S \cap \tilde{S} \neq \{0\}$$

de donde existe

$$z \in S \cap \tilde{S},\ z \neq 0$$

Entonces

$$\lambda_n < \frac{z^t\, Az}{z^t\, z} < \lambda_k$$

de donde

$$\min_{x \in S} \frac{x^t\, Ax}{x^t\, x} < \lambda_k$$

es decir

$$h_A(S) < \lambda_k$$

$$\text{con } S \in F_k$$

Tomemos ahora

$$S' = [x_1, x_2, \ldots, x_k]$$

si $x \in S'$, $x \neq 0$, tenemos

$$\lambda_k < \frac{x^t Ax}{x^t x} < \lambda_1$$

entonces

$$\min_{x \in S'} \frac{x^t Ax}{x^t x} > \lambda_k$$

es decir

$$h_\lambda(S') > \lambda_k$$

de donde

$$\max h_\lambda(S) = \lambda_k$$

$$\text{con} \quad S \in F_k \qquad \blacksquare$$

Observaremos a continuación que la importancia de este Teorema consiste en que puede establecer una desigualdad para determinados valores característicos de una matriz sin tener conocimientos previos sobre los valores y vectores característicos de esa matriz.

Es decir, obtenemos un cálculo aproximado para una λ_k en forma independiente de los demás valores y vectores característicos.

Para continuar con resultados acerca de perturbaciones en matrices simétricas requerimos del Lema siguiente:

<u>Lema.</u>

Sean A, B y E matrices simétricas tales que

$$A + E = B$$

y sean las funciones

$$H_A(S),\ H_E(S) \text{ y } H_B(S)$$

como se definieron antes.

Entonces

$$H_B(S) \leq H_A(S) + H_E(S)$$

para todo $S \in F_k$

<u>Demostración.</u>

Sean $$A + E = B$$

al multiplicar por $x \in S$ por la derecha y por la izquierda y al dividir entre $x^t x$ se tiene

$$\frac{x^t Ax}{x^t x} + \frac{x^t Ex}{x^t x} = \frac{x^t Bx}{x^t x}$$

pero

$$\max_{x \in S} \left(\frac{x^t Ax}{x^t x} + \frac{x^t Ex}{x^t x}\right) = \max_{x \in S} \left(\frac{x^t Bx}{x^t x}\right)$$

de donde

$$\max_{x \in S} \frac{x^t Bx}{x^t x} \leq \max_{x \in S} \frac{x^t Ax}{x^t x} + \max_{x \in S} \frac{x^t Ex}{x^t x}$$

es decir

$$H_B(S) \leq H_A(S) + H_E(S)$$

para toda $S \in F_k$

Con estos elementos, presentamos el siguiente resultado de gran importancia en la Teoría de Perturbación para matrices simétricas.

Teorema.

Sean A,B y E matrices simétricas tales que

$$A + E = B$$

cuyos valores característicos correspondientes son

$$\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$$

$$\beta_1 \geq \beta_2 \geq \ldots \geq \beta_n$$

$$\varepsilon_1 \geq \varepsilon_2 \geq \ldots \geq \varepsilon_n$$

entonces

$$\alpha_k + \varepsilon_n \leq \beta_k \leq \alpha_k + \varepsilon_1$$

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

Demostración.

Por el Lema anterior, tenemos

$$H_B(S) \leq H_A(S) + H_E(S)$$

para toda $S \in F_{n-k+1}$

Sean

$$x_1, x_2, \ldots, x_n$$

los vectores característicos ortonormales de A correspondien tes a

$$\alpha_1, \alpha_2, \ldots, \alpha_n$$

y sea además

$$S' = [x_k, x_{k+1}, \ldots, x_n]$$

Entonces, por el Teorema Minimax, tenemos

$$\min_{S \in F_{n-k+1}} H_A(S) = H_A(S') = \alpha_k$$

y como

$$\beta_k \leq H_B(S)$$

y

$$H_E(S) \leq \varepsilon_1$$

para toda $S \in F_{n-k+1}$

en particular

$$\beta_k \leq H_B(S')$$

y

$$H_E(S') \leq \varepsilon_1$$

entonces

$$H_B(S') \leq H_A(S') + H_E(S')$$

de donde se obtiene

$$\beta_k \leq H_B(S') \leq H_A(S') + H_E(S')$$

$$\leq \alpha_k + H_E(S')$$

$$\leq \alpha_k + \varepsilon_1$$

de donde

$$\beta_k \leq \alpha_k + \varepsilon_1$$

Ahora, si tomamos

$$C = - E$$

como

$$A = B - E$$

tenemos

$$A = B + C$$

y por el resultado probado antes, tenemos

$$\alpha_k < \beta_k + \gamma_1$$

donde γ_1 es el valor característico más grande de C.

Pero como

$$\gamma_1 = - \varepsilon_n$$

resulta que

$$\alpha_k < \beta_k - \varepsilon_n$$

es decir

$$\alpha_k + \varepsilon_n < \beta_k$$

De donde

$$\alpha_k + \varepsilon_n < \beta_k < \alpha_k + \varepsilon_1 \qquad \blacksquare$$

En algunas aplicaciones, la matriz E es una matriz de perturbación pequeña, y por lo mismo ε_1 y ε_n son muy pequeñas. Cunado esto ocurre, el teorema anterior establece que los valores característicos de B están muy cerca a los de A.

De hecho, la importancia de este Teorema radica en que a cada valor característico de A, le asocia un único valor característico de B, que se encuentra en un intervalo bien definido cerca del valor característico de A.

PERTURBACIONES EN VALORES SINGULARES.

Con el Teorema de Perturbación para valores característicos que presentamos en la sección anterior será posible derivar el correspondiente para valores singulares. Para ello requerimos establecer una relación entre los valores singulares de la matriz A y los valores característicos de otra matriz relacionada de alguna manera con A.

Para definir a ésta última matriz comenzaremos por recordar que los valores singulares de A son las raíces cuadradas positivas de $A^t A$, sin embargo, como trabajaremos con matrices perturbadas no nos conviene trabajar con esta expresión por resultar muy complicada:

$$(A + E)^t (A + E)$$

Sin embargo podemos partir de la Descomposición Singular de A.

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

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^t$$

de donde tenemos

$$Av_i = \sigma_i u_i$$

$$A^t u_i = \sigma_i v_i$$

que en forma matricial se puede expresar como

$$\begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \sigma_i \begin{bmatrix} u_i \\ v_i \end{bmatrix}$$

y si

$$S = \begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix}$$

tenemos la expresión para la matriz relacionada con A que buscabamos.

El resultado siguiente establece una relación entre los valores singulares de la matriz A y los valores característicos de la matriz S.

<u>Teorema.</u>

Sea Amxn, $m \geq n$ y sea también $S(m+n)x(m+n)$, la matriz simétrica definida por

$$S = \begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix}$$

Si los valores singulares de A, son

$$\sigma_1, \sigma_2, \ldots, \sigma_n$$

entonces, los valores característicos de S son

$$\sigma_1, \sigma_2, \ldots, \sigma_n, -\sigma_1, -\sigma_2, \ldots, -\sigma_n$$

y los $(m-n)$ restantes son ceros.

Con todos estos elementos, estamos ya en condiciones de presentar el teorema que se refiere a la Perturbación de los Valores Singulares de una matriz Amxn.

Teorema.

Sean A,B y E matrices de $m \times n$ tales que

$$B = A + E$$

Si denotamos sus respectivos valores singulares por

$$\beta_i,\ \alpha_i \text{ y } \varepsilon_i$$

$$\text{para } i = 1,2,\ldots,n$$

y cada conjunto ordenado en forma decreciente, entonces

$$|\beta_i - \alpha_i| \leq \varepsilon_1 \equiv \|E\|$$

$$\text{para } i = 1,2,\ldots,n\ .$$

Demostración.

Sean

$$\tilde{B} = \begin{bmatrix} 0 & B \\ B^t & 0 \end{bmatrix};\ \tilde{A} = \begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix};\ \tilde{E} = \begin{bmatrix} 0 & E \\ E^t & 0 \end{bmatrix}$$

entonces

$$\tilde{B} - \tilde{A} = \tilde{E}$$

y los valores característicos correspondientes son:

$$\{\tilde{\beta}_i\} = \{\beta_1,\beta_2,\ldots,\beta_n,\underbrace{0,\ldots,0}_{m-n} - \beta_n,-\beta_{n-1},\ldots,-\beta_1\}$$

$$\{\tilde{\alpha}_i\} = \{\alpha_1,\alpha_2,\ldots,\alpha_n,\underbrace{0,\ldots,0}_{m-n} - \alpha_n,-\alpha_{n-1},\ldots,-\alpha_1\}$$

$$\{\tilde{\varepsilon}_i\} = \{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n,\underbrace{0,\ldots,0}_{m-n} - \varepsilon_n,-\varepsilon_{n-1},\ldots,-\varepsilon_1\}.$$

Pero como por el Teorema anterior, para matrices simé-tricas tenemos que

$$\tilde{\alpha}_i + \tilde{\varepsilon}_{n+m} < \tilde{\beta}_i < \tilde{\alpha}_i + \tilde{\varepsilon}_1$$

para $i = 1,2,\dots,(n+m)$

al ponér la expresión anterior en términos de α_i, β_i y ε_i

para $i = 1,2,\dots,n$

se obtiene

$$\alpha_i - \varepsilon_1 < \beta_i < \alpha_i + \varepsilon_1$$

para $i = 1,2,\dots,n$

de donde $\quad -\varepsilon_1 < \beta_i - \alpha_i < \varepsilon_1 \quad$ para $i = 1,2,\dots,n.$

es decir $\quad |\beta_i - \alpha_i| < \varepsilon_1 \equiv \|E\| \quad$ para $i = 1,2,\dots,n.$

Pasamos ahora a dar condiciones para que la matriz

$$\tilde{A} = A + E$$

tenga una seudoinversa $\tilde{A}^+$ que esté cercana a A^+.

La primera condición necesaria es la conservación del rango. La motivación de este concepto se presentó al inicio de este capítulo por medio de un ejemplo en el que se hizo ver que si

$$\text{rango } (\tilde{A}) > \text{rango } (A)$$

entonces es posible que

$$\|\tilde{A}^+ - A^+\|$$

sea muy grande, aún cuando $\|E\|$ sea muy pequeña.

Esta situación no se presentó en el caso cuadrado, cuando $m = n$, porque la matriz A tenía rango máximo. Y tampoco ocurre cuando A_{mxn}, $m > n$, tiene rango $(A) = n$, es decir, cuando A tiene rango máximo.

De lo anterior se deduce que se presentan problemas cuando

$$\text{rango } (A) < n$$

es decir, cuando A es de rango deficiente. Pues es entonces cuando existen perturbaciones E muy pequeñas tales que

$$\| \tilde{A}^{+} - A^{+} \|$$

es muy grande.

<u>Teorema.</u>

Si rango $(A) = k < n$ y

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

es la descomposición singular de A, al tomar a E como

$$E = U \begin{bmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \varepsilon & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} V^t \qquad \text{con } \varepsilon > 0$$

$$k+1$$

tal que $\tilde{A} = A + E$

$$= U \begin{bmatrix} \sigma_1 & & & & & & & \\ & \sigma_2 & & & & & 0 & \\ & & \ddots & & & & & \\ & & & \sigma_k & & & & \\ & & & & \varepsilon & & & \\ & & & & & 0 & & \\ & 0 & & & & & \ddots & \\ & & & & & & & 0 \end{bmatrix} V^t$$

entonces

$$\|E\| = \varepsilon,$$

$$\text{rango } (\tilde{A}) = k+1$$

y

$$\|\tilde{A}^+ - A^+\| = \frac{1}{\varepsilon}$$

Ahora bien, para conservar el rango es necesario conocer lo que se debe evitar para que éste disminuya. Y por lo anterior, es de esperar que

$$\|E\| < \sigma_k$$

para que el rango no disminuya.

Así pues tenemos el siguiente resultado que formaliza lo anterior:

<u>Teorema.</u>

Si E es tal que

$$\tilde{A} = A + E$$

$$\text{rango } (\tilde{A}) < \text{rango } (A) = k$$

entonces

$$\|E\| = \varepsilon_1 > \sigma_k$$

donde ε_1 es el valor singular más grande de E y σ_k el valor singular positivo más pequeño de A.

Demostración.

Sean

$$\{\sigma_i\}, \ \{\tilde{\sigma}_i\} \ \text{ y } \ \{\varepsilon_i\}$$

los valores singulares de A, $\tilde{A}$ y E respectivamente.

Por el teorema de perturbación de valores singulares tenemos

$$|\tilde{\sigma}_i - \sigma_i| < \varepsilon_1 = \|E\|$$

de donde

$$\tilde{\sigma}_k - \sigma_k < \varepsilon_1$$

$$\text{y} \qquad \sigma_k - \tilde{\sigma}_k < \varepsilon_1$$

y como por hipótesis

$$\tilde{\sigma}_k = 0$$

tenemos de la última desigualdad que

$$\sigma_k < \varepsilon_1 \quad \blacksquare$$

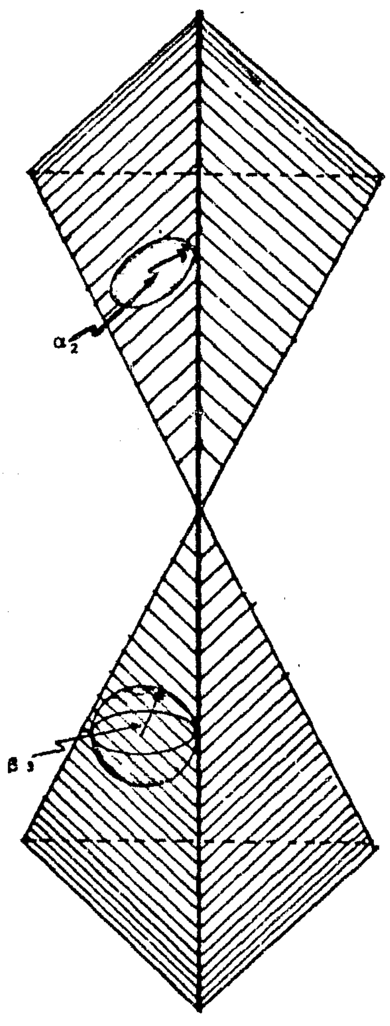
Lo anterior se resume en el siguiente:

Teorema.

Si rango $(A) = k$, entonces

$$\sigma_k = \min\{\|A - B\| / \text{rango } (B) < \text{rango } (A)\},$$

Para mejor comprensión del resultado anterior consideramos la siguiente figura:



En la figura se puede ver un cuerpo geométrico en $\mathbb{R}^3$ que representa algunos subespacios de las matrices de 3 x 3. De este modo, el espacio delimitado por las caras contiene a las matrices de 3 x 3 con rango 3, los planos del cuerpo contienen a las matrices de 3 x 3 de rango 2, las aristas, a las matrices de 3 x 3 con rango 1 y el vértice a la matriz de rango cero.

Entonces, el círculo sobre una de las caras, representa a una matriz A 3 x 3 con rango (A) = 2, y α_2, su valor singular, más pequeño, es la distancia mínima del círculo a la arista más cercana.

La esfera, dentro del cuerpo, representa a una matriz B 3x3 con rango (B) = 3, y β_3 su valor singular más pequeño, es la distancia mínima de la esfera al plano más cercano.

Hasta este momento hemos obtenido ya dos condiciones para que $\tilde{A}^+$ esté cerca de A^+:

Necesitamos que el rango de A no aumente y es necesario también que no disminuya.

Esto nos lo confirma el siguiente:

Teorema.

Sea
$$\tilde{A} = A + E$$

tal que

i) rango $(\tilde{A}) <$ rango (A)

ii) $\|E\| = \varepsilon_1 < \sigma_k$

entonces
$$\text{rango }(\tilde{A}) = \text{rango }(A)$$

y
$$\frac{1}{\tilde{\sigma}_k} = \|\tilde{A}\| < \frac{\|A^+\|}{1 - \|E\| \cdot \|A^+\|} = \frac{1}{\sigma_k - \varepsilon_1}$$

Demostración.

Sean
$$\{\tilde{\sigma}_i\},\ \{\sigma_i\} \text{ y } \{\varepsilon_i\}$$

los valores singulares de $\tilde{A}$, A y E.

Por hipótesis
$$\tilde{\sigma}_{k+1} = \tilde{\sigma}_{k+2} = \ldots = \tilde{\sigma}_n = 0$$

Se probará que
$$\sigma_k \neq 0.$$

Tenemos pues que
$$\tilde{A} - A = E$$

entonces por el teorema de perturbación de valores singulares
$$|\tilde{\sigma}_i - \sigma_i| < \varepsilon_1$$

al tomar $i = k$, resulta

$$|\tilde{\sigma}_k - \sigma_k| < \varepsilon_1$$

entonces

$$|\sigma_k - \tilde{\sigma}_k| < \varepsilon_1$$

es decir

$$\sigma_k - \tilde{\sigma}_k < \varepsilon_1$$

o también

$$0 < (\sigma_k - \varepsilon_1) < \tilde{\sigma}_k \neq 0$$

de donde

$$\text{rango } (\tilde{A}) = k$$

y además

$$\tilde{\sigma}_k > (\sigma_k - \varepsilon_1) > 0$$

por lo tanto

$$\frac{1}{\tilde{\sigma}_k} < \frac{1}{\sigma_k - \varepsilon_1} \quad .$$

∎

<u>CONTINUIDAD Y DIFERENCIABILIDAD DE SEUDOINVERSAS.</u>

Los resultados anteriores sólo nos han permitido estimar la norma de $(A + E)^+$:

$$\| (A + E)^+ \|$$

y saber que ésta tiende a la norma de A^+:

$$\| A^+ \|$$

bajo ciertas condiciones.

Sin embargo, nos hace falta estimar

$$\| (A+E)^{+} - A^{+} \|$$

y hacer ver que $(A+E)^{+}$ tiende a A^{+} cuando E tiende a cero.

Para calcular la norma de esta diferencia trataremos de obtener una expresión para

$$(A+E)^{+} - A^{+}$$

en función de

$$(A+E) - A$$

Comenzaremos por considerar el caso cuando

$$A^{+} = A^{-1}$$

ya que es fácil observar lo siguiente:

$$\begin{aligned} (A+E)^{-1} - A^{-1} &= (A+E)^{-1} [I - (A+E) A^{-1}] \\ &= (A+E)^{-1} [A - (A+E)] A^{-1} \\ &= - (A+E)^{-1} [(A+E) - A] A^{-1} \end{aligned}$$

De lo anterior, parece natural proponer la siguiente expresión para la diferencia de Seudoinversas. Por comodidad haremos

$$B = (A+E)$$

entonces

$$B^{+} - A^{+} = - B^{+} (B - A) A^{+} + C (B,A)$$

donde C(B,A) es una corrección a la expresión y que trataremos de obtener para el caso general.

Para calcular el valor de dicha corrección, simplemente despejamos:

$$C(B,A) = B^+ - A^+ + B^+(B - A)\ A^+$$
$$= B^+ - A^+ + B^+\ BA^+ - B^+\ AA^+$$

de donde

$$C(B,A) = B^+(I - AA^+) - (I - B^+\ B)\ A^+$$

al sustituir el valor de $C(B,A)$, obtenemos la siguiente expresión:

$$B^+ - A^+ = - B^+(B - A)\ A^+ + B^+(I - AA^+) - (I - B^+B)\ A^+$$

que viene a ser una primera fórmula para la diferencia $B^+ - A^+$

Pero como podemos ver, la diferencia $B - A$ no aparece en todos los términos y para nuestros propósitos es necesario que aparezca. Así que trataremos de obtener una expresión para cada término en función de $B - A$.

Para

$$B^+(I - A\ A^+)$$

tenemos que

$$B^+(I - A\ A^+) = B^+\ (I - P_{\mathcal{R}(A)})$$

es decir

$$B^+\ (I - A\ A^+) = B^+\ P_{\mathcal{R}(A)^\perp}$$

La clave para obtener lo que deseamos está en buscar una representación adecuada para

$$P_{\mathcal{R}(A)^\perp} = P_{N(A^t)}$$

y en utilizar la propiedad

$$B^+ = B^+\ P_{\mathcal{R}(B)}$$

Así pues, tenemos

$$B^+(I - A\ A^+) = B^+\ P_{R(B)} P_{N(A^t)}$$

Pero, si observamos $P_{R(B)}$ también puede expresarse como

$$P_{R(B)} = P_{N(B^t)^\perp} = (B^t)^+\ B^t$$

con lo que obtenemos

$$B^+(I - A\ A^+) = B^+(B^t)^+ B^t P_{N(A^t)}$$

Y para obtener una expresión en términos de

$$B - A$$

agregamos un término nulo de tal forma que se pueda sacar como factor a

$$B - A$$

Así pues, obsérvese que

$$A^t\ P_{N(A^t)} = 0$$

y por lo tanto

$$F A^t\ P_{N(A^t)} = 0$$

con F arbitraria.

De donde

$$B^+(I - A A^+) = B^+(B^t)^+\ B^t P_{N(A^t)} + F\ A^t\ P_{N(A^t)}$$

y así tenemos

$$B^+(I - A\ A^+) = [B^+\ (B^t)^+ B^t - F\ A^t]\ P_{N(A^t)}$$

al hacer $$F = B^+ (B^t)^+$$

se obtiene
$$B^+ (I - A\,A^+) = B^+ (B^t)^+ (B - A)^t (I - A\,A^+)$$

Para el término
$$-(I - B^+ B)\,A^+$$
ocurre algo análogo:

Se tiene que
$$-(I - B^+ B) A^+ = -\ (I - P_{N(B)^{\perp}})\ A^+$$
$$= -\ P_{N(B)}\ A^+$$

Y como
$$P_{N(B)^{\perp}} = P_{\mathcal{R}(B^t)}$$
tenemos que
$$P_{N(B)}\ B^t\ F = 0$$
para cualquier F.

Entonces
$$-\ (I - B^+ B) A^+ = -\ P_{N(B)} A^+ + P_{N(B)} B^t\ F$$
$$=\ \ P_{N(B)} (B^t\ F - A^+)$$

pero
$$A^+ = P_{\mathcal{R}(A^t)}\ A^+$$
$$= A^t\ (A^t)^+\ A^+$$

Así que
$$-\ (I - B^+ B)\ A^+ = P_{N(B)} [\,B^t F - A^t (A^t)^+\ A^+]$$

y al tomar
$$F = (A^t)^+ A^+$$

se obtiene

$$- (I - B^+ B) A^+ = P_{N(B)} (B - A)^t (A^t)^+ A^+$$

$$= (I - B^+ B) (B - A)^t (A^t)^+ A^+$$

Con esto podemos ya expresar a la corrección

$$C(B, A)$$

en términos de la diferencia $(B - A)$:

$$C(B,A) = B^+(I - A A^+) - (I - B^+ B) A^+$$

$$= B^+(B^t)^+ (B - A)^t (I - A A^+) +$$

$$(I - B^+B) (B - A)^t (A^t)^+ A^+$$

Lo anterior nos permite enunciar el resultado formalmente como un Teorema.

<u>Teorema.</u>

La matriz $(B^+ - A^+)$

satisface

$$(B^+ - A^+) = - B^+(B-A)A^+ + B^+(I - P_{\mathcal{R}(A)})$$

$$- (I - P_{N(B)^\perp}) A^+$$

y haciendo notar la dependencia de

$$(B^{+} - A^{+}) \text{ en } (B - A)$$

tenemos la siguiente expresión alternativa:

$$\begin{aligned}(B^{+} - A^{+}) = &- B^{+} (B - A) A^{+} \\ &+ B^{+}(B^{t})^{+} (B - A)^{t} (I - A A^{+}) \\ &+ (I - B^{+}B) (B - A)^{t} (A^{t})^{+} A^{+}.\end{aligned}$$

Conviene hacer notar que el término

$$- B^{+}(B - A) A^{+}$$

resulta una expresión suficiente para

$$(B^{+} - A^{+})$$

en el caso de que la matriz fuera cuadrada e invertible. Mientras que la existencia de los términos restantes está relacionada con los casos cuando la matriz no es cuadrada o es de rango deficiente.

Pues

$$B^{+}(B^{t})^{+} (B - A)^{t} (I - A A^{+})$$

es diferente de cero sólo si

$$\text{rango } (A) < m$$

y

$$(I - B^{+} B) (B - A)^{t} (A^{t})^{+} A^{+}$$

es diferente de cero sólo si

$$\text{rango } (B) < n$$

ya que

$$A\,A^{+} = I \quad \text{si rango } (A) = m \quad y$$

$$B^{+}\,B = I \quad \text{si rango } (B) = n$$

Como consecuencia de la fórmula anterior y los resultados que le preceden obtenemos:

Teorema.

Sea
$$\tilde{A} = A + E$$

tal que
$$\text{rango } (\tilde{A}) \leq \text{rango } (A)$$

y si $E \to 0$,

entonces
$$\|\tilde{A}^{+} - A^{+}\| \to 0$$

es decir
$$\lim_{E \to 0} \tilde{A}^{+} = A^{+} .$$

Su demostración es inmediata a partir de la fórmula anterior haciendo

$$B = \tilde{A}$$

y observando
$$\|\tilde{A}^{+}\| \to \|A^{+}\|$$

cuando $$(B - A) = (\tilde{A} - A) \to 0.$$

También podemos interpretar el resultado anterior como sigue:

Si variamos un poco los elementos de la matriz A, de tal forma que la matriz obtenida $\tilde{A}$ no tenga rango mayor que el de A, entonces $\tilde{A}^+$ está cerca de A^+, es decir, A^+ es una función continua de A siempre y cuando A no cambie de rango.

En particular, podemos considerar una matriz $A(\varepsilon)$ cuyos elementos dependan del parámetro real ε en forma continua o diferenciable.

Con esto, es natural preguntarnos si $A^+(\varepsilon)$ es entonces una función continua o diferenciable de ε.

El resultado siguiente nos da condiciones para que esto suceda.

Teorema.

Si rango $(A(\varepsilon)) \leq$ rango (A)

entonces

I) $A(\varepsilon)$ continua en $\varepsilon = 0$

implica

$A^+(\varepsilon)$ es continua en $\varepsilon = 0$

II) $A(\varepsilon)$ diferenciable en $\varepsilon = 0$

implica

$A^+(\varepsilon)$ es diferenciable en $\varepsilon = 0$

y además:

$$\frac{d\,A^{+}(\varepsilon)}{d\,\varepsilon} = -A^{+}(\varepsilon)\frac{d\,A(\varepsilon)}{d\,\varepsilon}A^{+}(\varepsilon) +$$

$$+A^{+}(\varepsilon)(A^{t}(\varepsilon))^{+}\left(\frac{d\,A(\varepsilon)}{d\,\varepsilon}\right)^{t}(I - P_{R(A(\varepsilon))}) +$$

$$-(I - P_{N(A(\varepsilon))^{\perp}})\left(\frac{d\,A(\varepsilon)}{d\,\varepsilon}\right)^{t}(A^{t}(\varepsilon))^{+}A^{+}(\varepsilon)$$

La demostración de este resultado es también inmediata a partir de la fórmula para la diferencia de seudoinversas. La omitimos por considerarla rutinaria.

TEORIA DE PERTURBACION DE PRIMER ORDEN.

A estas alturas estamos ya en condiciones de obtener una primera estimación del efecto que tienen las perturbaciones sobre la solución del problema de Cuadrados Mínimos.

Consideremos entonces la solución x de norma mínima del problema

$$\min_{z} \|b - Az\|^{2} = \|b - Ax\|^{2}$$

Como sabemos, x está dada por

$$x = A^{+}b$$

Supongamos que A y b se perturban de la siguiente manera:

$$A(\varepsilon) = A + \varepsilon E$$

$$b(\varepsilon) = b + \varepsilon e$$

donde ε es un número real y pequeño, E y e son una matriz y un vector de perturbación respectivamente, tales que

$$\text{rango}\ (A(\varepsilon)) = \text{rango}\ (A)$$

y $A(\varepsilon)$ es una función diferenciable de ε como se ve, con $A(0) = A$.

De acuerdo con lo anterior, es natural considerar para cada ε a la solución $x(\varepsilon)$ de norma mínima del problema.

$$\min_z \| b(\varepsilon) - A(\varepsilon) z \|^2 = \| b(\varepsilon) - A(\varepsilon)\, x(\varepsilon) \|^2$$

con

$$x(0) = x$$

Así como esperar que $x(\varepsilon)$ esté "cerca de" x cuando ε es pequeña.

A continuación trataremos de verificar esto a partir de la expresión.

$$x(\varepsilon) = A^+(\varepsilon)\, b(\varepsilon)$$

El desarrollo de Taylor de esta expresión alrededor del origen conduce a lo siguiente:

$$x(\varepsilon) = x(0) + \left[\frac{d}{d\varepsilon} (A^+(\varepsilon) \cdot b(\varepsilon)) \right]_{\varepsilon=0} \cdot \varepsilon + O(\varepsilon^2)$$

en donde

$$\frac{d}{d\varepsilon}\,(A^{+}(\varepsilon)\cdot b(\varepsilon)) = \left[\frac{d}{d\varepsilon}\,A^{+}(\varepsilon)\right]\cdot b(\varepsilon) + A^{+}(\varepsilon)\cdot e$$

y por el último teorema de la sección anterior, tenemos

$$\left[(\frac{d}{d\varepsilon}\,A^{+}(\varepsilon))b(\varepsilon)\right]_{\varepsilon=0} = -\,A^{+}\ E\ A^{+}b\ +$$

$$+\ A^{+}(A^{t})^{+}\ E\ [\,I - P_{\mathcal{R}(A)}]\ b +$$

$$-\ [\,I - P_{N(A)^{\perp}}]\ E\ (A^{t})^{+}A^{+}b + A^{+}\,e.$$

De donde

$$(x(\varepsilon) - x) = \varepsilon\left\{-A^{+}\ Ex + [\,A^{+}(A^{t})^{+}E(I - P_{\mathcal{R}(A)})\ b\ +\right.$$

$$\left. +(I - P_{N(A)^{\perp}})E(A^{t})^{+}A^{+}]\,b + A^{+}\,e\right\} + O(\varepsilon^{2})$$

y al simplificar resulta

$$(x(\varepsilon) - x) = \varepsilon\left\{-A^{+}\ Ex + A^{+}(A^{t})^{+}Er + P_{N(A)}E(A^{t})^{+}\ x\ +\right.$$

$$\left. A^{+}\ e\right\} + O\ (\varepsilon^{2})$$

Al aplicar normas

$$\|x(\varepsilon) - x\| \leq \varepsilon\left\{\|A^{+}\|\cdot\|A\|\cdot\frac{\|E\|}{\|A\|}\cdot\|x\|\ +\right.$$

$$+\ \|A\|^{2}\cdot\|A^{+}\|^{2}\cdot\frac{\|E\|}{\|A\|}\cdot\frac{\|r\|}{\|A\|\cdot\|x\|}\cdot\|x\|\ +$$

$$+ \frac{\|E\|}{\|A\|} \cdot \|A\|^{+} \cdot \|A\| \cdot \|x\| +$$

$$\left. + \|A^{+}\| \cdot \|A\| \cdot \frac{\|e\|}{\|A\|} \cdot \frac{\|x\|}{\|x\|} \right\} + O\,(\varepsilon^{2})$$

tomando

$$k(A) = \|A^{+}\| \cdot \|A\|$$

$$\rho_A = \frac{\|E\|}{\|A\|}$$

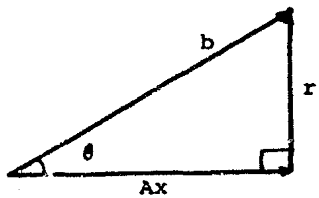
$$\rho_b = \frac{\|e\|}{\|b\|}$$

tenemos

$$\frac{\|x(\varepsilon) - x\|}{\|x\|} \leq \varepsilon\ k(A) \left\{ \left(2 + k(A)\ \frac{\|r\|}{\|A\| \cdot \|x\|}\right) \rho_A + \right.$$

$$\left. + \frac{\|b\|}{\|A\| \cdot \|x\|}\ \rho_b \right\} + O\,(\varepsilon^{2})$$

Esta última expresión puede escribirse en forma muy simple al introducir el ángulo θ que forman b y Ax, véase la figura siguiente



ya que

$$\frac{\|b\|}{\|Ax\|} = \frac{1}{\cos\,\theta}$$

$$\frac{\|r\|}{\|A\|\cdot\|x\|} < \frac{\|r\|}{\|Ax\|} = \tan\theta$$

$$\frac{\|b\|}{\|A\|\cdot\|x\|} < \frac{\|b\|}{\|Ax\|} = \frac{1}{\cos\theta}$$

Por lo tanto,

$$\frac{\|x(\varepsilon)-x\|}{\|x\|} < \varepsilon\left[\frac{k(A)}{\cos\theta}\right]\left\{(2\cos\theta + k(A)\ \mathrm{sen}\,\theta)\,\rho_A + \right.$$

$$\left. + \rho_b\right\} + O\ (\varepsilon^2)$$

y como

$$\frac{-\pi}{2} < \theta < \frac{\pi}{2}$$

$$2\cos\theta < 2$$

entonces

$$\frac{\|x(\varepsilon)-x\|}{\|x\|} < \varepsilon\left[\frac{k(A)}{\cos\theta}\right]\left\{(2 + k(A)\ \mathrm{sen}\ \theta)\ \rho_A + \rho_b\right\} +$$

$$+ O\ (\varepsilon^2)$$

Esta última expresión nos indica que si el ángulo θ entre b y Ax es muy pequeño, el factor k(A) es el que puede amplificar los errores relativos ρ_A y ρ_b. Pero si el ángulo θ no es pequeño es decir, si $\|r\|$ es grande, entonces el factor de amplificación puede ser $k(A)^2$.

En la práctica, si $\|r\|$ y $k(A)$ son grandes, esto equivale a tener un problema con una solución muy sensible a pequeñas alteraciones.

CAPITULO V

SIGNIFICADO GEOMETRICO DE LAS COTAS DE PERTURBACION

INTRODUCCION.

Como ya se vió, la solución al problema de encontrar la $\tilde{x}$ de norma mínima que minimiza

$$\| b - Ax \|^2$$

la podemos obtener de la siguiente forma:

i) Calcular la proyección de b sobre el $\mathcal{R}(A)$, es decir, obtener

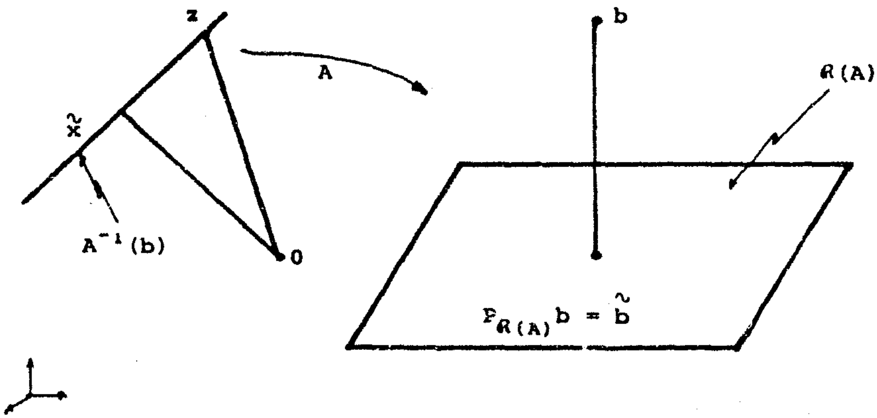
$$P_{\mathcal{R}(A)} b = \tilde{b}$$

ii) Encontrar una solución z de

$$Az = \tilde{b}$$

iii) Calcular la proyección de z sobre $N(A)^{\perp}$ para encontrar la $\tilde{x}$ buscada.

Los tres pasos anteriores los ilustra la siguiente figura:



Cuando se perturba A, es natural que cada uno de los subproblemas anteriores quede perturbado y también que cada uno de ellos contribuya con un término en el error global. A continuación se verá que las variaciones producidas por las perturbaciones se pueden interpretar geométricamente, lo cual nos facilitará su cálculo.

Consideremos ahora el siguiente caso particular:

¿Cómo varía la proyección ortogonal de y sobre el subespacio S si se perturba dicho subespacio S?.

Es decir, si S_1 y S_2 son dos subespacios "cercanos" ¿qué tanto pueden diferenciarse las correspondientes proyecciones ortogonales de y?.

Si P_1 es la proyección ortogonal sobre S_1 y P_2 la correspondiente sobre S_2, tenemos

$$\begin{aligned} \|P_1y - P_2y\| &= \|(P_1 - P_2)\,y\| \\ &\leq \|P_1 - P_2\| \cdot \|y\| \\ &\leq 2 \cdot \|y\| \end{aligned}$$

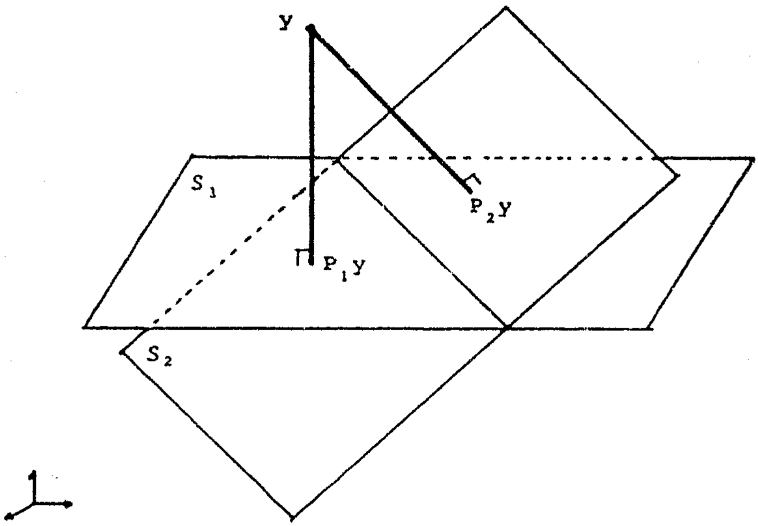
ya que

$$\|P_1\| = \|P_2\| = 1$$

pero esta cota no nos es muy útil ya que no manifiesta si S_1 y S_2 son "cercanos".

Grádicamente la situación si S_1 y S_2 son dos planos en

$\mathbb{R}^3$, es:



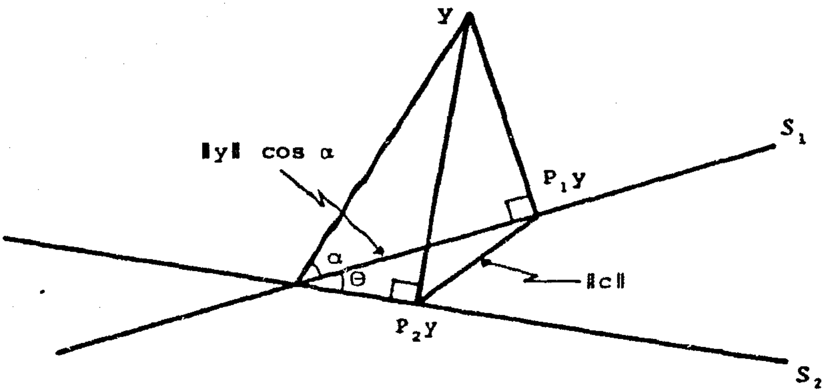
Para comprender mejor el problema consideraremos primero el caso cuando S_1 y S_2 son subespacios de dimensión 1, y y el vector que se va a proyectar:

Sean entonces

$$S_1,\ S_2 \subset \mathbb{R}^2, \text{ tales que}$$

$$\dim(S_1) = \dim(S_2) = 1$$

como se ven en la figura:



donde $c = P_1y - P_2y$

Obsérvese que por la Ley de los Senos

$$\frac{\|c\|}{\text{sen } \theta} = \frac{\|y\| \cos \alpha}{\text{sen } (90° + \alpha)}$$

pero

$$\text{sen}(90° + \alpha) = \cos \alpha$$

de donde

$$\frac{\|c\|}{\text{sen } \theta} = \|y\|$$

por lo tanto

$$\|c\| = \|y\| \text{ sen } \theta$$

De aquí podemos concluir que si S_1 y S_2 son dos subespacios en $\mathbb{R}^2$ de dimensión 1 el problema resulta ser muy sencillo pues entonces

$$\|P_1y - P_2y\| = \|y\| \cdot |\text{sen } \theta|$$

donde θ es el ángulo formado por S_1 y S_2. Además, podemos interpretar que el hecho de que los subespacios estén "cercanos" significa que el ángulo θ sea pequeño o que $|\operatorname{sen}\theta|$ sea pequeño.

El caso anterior nos permite considerar el caso cuando S_1 y S_2 son subespacios en $\mathbb{R}^n$ de dimensión 1, veamos que sucede:

Sea
$$\mathbb{R}^n = (S_1 + S_2) \oplus (S_1 + S_2)^{\perp}$$

y sea
$$y = \tilde{y} + \tilde{\tilde{y}}$$

con
$$\tilde{y} \in (S_1 + S_2)$$
$$\tilde{\tilde{y}} \in (S_1 + S_2)^{\perp}$$

tenemos entonces
$$P_1 y = P_1 \tilde{y} + P_1 \tilde{\tilde{y}}$$
$$= P_1 \tilde{y}$$
$$P_2 y = P_2 \tilde{y} + P_2 \tilde{\tilde{y}}$$
$$= P_2 \tilde{y}$$

De donde
$$P_1 y - P_2 y = P_1 \tilde{y} - P_2 \tilde{y}$$

y al aplicar lo obtenido para el caso $\mathbb{R}^2$ obtenemos

$$\| P_1 y - P_2 y \| = \| P_1 \tilde{y} - P_2 \tilde{y} \|$$
$$= \| \tilde{y} \| \cdot |\operatorname{sen}\theta|$$

es decir

$$\|P_1y - P_2y\| = \frac{\|\tilde{y}\| \cdot |\text{sen }\theta|}{\|y\|} \cdot \|y\|$$

de donde

$$\|P_1y - P_2y\| = k(y) \cdot \|y\|$$

con

$$k(y) = \frac{\|\tilde{y}\| \cdot |\text{sen }\theta|}{\|y\|} \leq |\text{sen }\theta| < 1.$$

Este desarrollo da lugar a la siguiente conclusión:

Teorema.

Si

$$S_1 \text{ y } S_2 \subset \mathbb{R}^n \text{ y}$$

$$\dim(S_1) = \dim(S_2) = 1$$

tenemos

$$\|P_1y - P_2y\| = k(y) \cdot \|y\|$$

donde

$$k(y) = \frac{\|\tilde{y}\| \cdot |\text{sen }\theta|}{\|y\|}$$

y entonces

$$\|P_1 - P_2\| < 1.$$

Conviene notar que la expresión

$$\|P_1 - P_2\| < 1$$

parece depender del seno del "ángulo" entre los subespacios S_1 y S_2 .

Pero... ¿será esto cierto en general?

La respuesta a esta pregunta la responderemos en la sección siguiente, donde presentaremos también el significado geométrico de la expresión

$$\| P_1 - P_2 \| < 1$$

Sólo nos queda hacer notar que el cálculo detallado y preciso de

$$\| P_1 - P_2 \|$$

será de gran utilidad para establecer las cotas de perturbación para la solución del problema de Cuadrados Mínimos.

CASO NO AGUDO: $\| P_1 - P_2 \| = 1$

En la sección anterior acabamos de ver que

$$\| P_1 - P_2 \| < 1$$

en ciertos casos particulares debido a que la norma de la diferencia $(P_1 - P_2)$ estaba acotada por el seno del "ángulo" que formaban los subespacios S_1 y S_2.

Comenzaremos por establecer

$$\| P_1 - P_2 \| < 1$$

en general, independientemente de las dimensiones de los subespacios S_1 y S_2. Esta vez, no desarrollaremos los conceptos de "ángulos" entre subespacios, sino que recurriremos a las propiedades de las proyecciones ortogonales por ser éstas menos elaboradas que los primeros.

Así, pues, el siguiente resultado establece la validez de nuestra expresión en general.

Teorema.

Sean P_1 y P_2 las proyecciones ortogonales sobre los subespacios S_1 y S_2 respectivamente, entonces

$$\|P_1 - P_2\| \leq 1.$$

Demostración.

$$\|(P_1 - P_2)x\|^2 = [(P_1 - P_2)x]^t [(P_1 - P_2)x]$$

$$= x^t (P_1 - P_2)^t (P_1 - P_2) x$$

y como

$$P_1^t = P_1 \quad y \quad P_2^t = P_2$$

tenemos

$$\|(P_1 - P_2)x\|^2 = x^t (P_1 - P_2)^2 x$$

Falta ahora hacer ver que

$$x^t (P_1 - P_2)^2 x \leq \|x\|^2$$

La idea para demostrarlo consiste en expresar al vector x en términos del vector

$$(P_1 - P_2) x$$

de tal forma que se obtenga algo parecido a:

$$\|x\|^2 = \|(P_1 - P_2)x\|^2 + \|y\|^2$$

Sin embargo, no es necesario que la matriz identidad I sea la que se escriba en términos de $(P_1 - P_2)$, sino que puede ser cualquier transformación ortogonal T.

Sea pues

$$\begin{aligned} T &= P_1 - P_2 + P_2 - P_1^{\perp} \\ &= P_1 - (I - P_1) \\ &= 2P_1 - I \end{aligned}$$

ya que $(I - P_1) = P_1^{\perp}$

de donde se ve que T es una reflexión:

i) $T = T^t$

ii) $T^2 = (2P_1 - I)(2P_1 - I)$

$\quad = I$

Así que $\|Tx\|^2 = \|x\|^2$

pero

$$\begin{aligned} \|Tx\|^2 &= \|(P_1 - P_2)x + (P_2 - P_1^{\perp})x\|^2 \\ &= [(P_1-P_2)x + (P_2-P_1^{\perp})x]^t[(P_1-P_2)x + (P_2-P_1^{\perp})x] \\ &= x^t(P_1-P_2)^2x + x^t(P_2 - P_1^{\perp})^2x + \\ &+ x^t(P_1-P_2)(P_2-P_1^{\perp})x + x^t(P_2-P_1^{\perp})(P_1-P_2)x \end{aligned}$$

Como los dos primeros términos son no negativos falta sólo analizar a los dos últimos.

Para esto, los desarrollamos y comparamos:

$$\begin{aligned}(P_1-P_2)(P_2-P_1^{\perp}) &= P_1P_2 - P_1P_1^{\perp} - P_2^2 + P_2P_1^{\perp}\\ &= P_1P_2 - P_2 + P_2\ (I-P_1)\\ &= P_1P_2 - P_2P_1\end{aligned}$$

y

$$\begin{aligned}(P_2-P_1^{\perp})(P_1-P_2) &= P_2P_1 - P_2^2 - P_1^{\perp}P_1 + P_1^{\perp}\ P_2\\ &= P_2P_1 - P_2 + P_1^{\perp}P_2\\ &= P_2P_1 - P_2 + (I-P_1)\ P_2\\ &= P_2P_1 - P_1P_2\end{aligned}$$

y al comparar obtenemos que:

$$(P_1-P_2)(P_2-P_1^{\perp}) = -\ (P_2-P_1^{\perp})(P_1-P_2)$$

y por consiguiente la suma de los dos últimos términos en la expresión para $\|x\|^2$ se anulan, por lo que obtenemos

$$\begin{aligned}\|Tx\|^2 &= \|(P_1-P_2)x\|^2 + \|(P_2-P_1^{\perp})x\|^2\\ &= \|x\|^2\end{aligned}$$

de esta igualdad obtenemos inmediatamente

$$\|P_1-P_2\| \leq 1.$$

Del resultado anterior, la primera pregunta que surge es:

¿Bajo qué condiciones sucede

$$\|P_1 - P_2\| = 1?$$

El hecho de que

$$\|P_1 - P_2\| = 1$$

quiere decir que existe $x \neq 0$, tal que

$$(P_1 - P_2)x = \pm x$$

Supongamos primero que

$$(P_1 - P_2)x = x$$

entonces

$$P_1x - P_2x = x$$

$$-P_2x = x - P_1x$$

sea

$$z = -P_2x$$

$$= P_1^{\perp}x$$

Si

$$z = 0,$$

entonces

$$x \in S_2^{\perp} \quad y \quad x \in S_1$$

es decir

$$x \in S_2^{\perp} \cap S_1$$

de donde

$$S_2^{\perp} \cap S_1 \neq \{0\}.$$

Si $$z \neq 0,$$

entonces $$z \in S_2 \quad y \quad z \in S_1^{\perp}$$

es decir
$$z \in S_2 \cap S_1^{\perp}$$

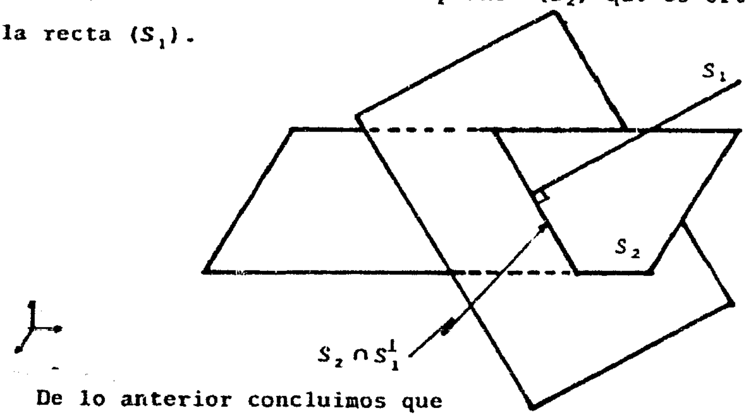
de donde
$$S_2 \cap S_1^{\perp} \neq \{0\}$$

El caso, cuando
$$(P_1 - P_2)x = -x$$

es similar al anterior

La siguiente figura en $\mathbb{R}^3$ ilustra la situación anterior, donde S_1 y S_2 son de dimensiones 1 y 2 respectivamente y
$$S_2 \cap S_1^{\perp} \neq \{0\}$$

es decir, existe un vector en el plano (S_2) que es ortogonal a la recta (S_1).



De lo anterior concluimos que

<u>Teorema.</u> $$\|P_1 - P_2\| = 1$$

si y sólo si

$$S_2^{\perp} \cap S_1 \neq \{0\} \text{ o } S_1^{\perp} \cap S_2 \neq \{0\}.$$

Este Teorema nos dice que

$$\|P_1 - P_2\| = 1$$

si y sólo si existe un vector en uno de los subespacios que es ortogonal a todos los vectores del otro subespacio.

Y el resultado siguiente nos indica bajo qué condiciones sucede que

$$\|P_1 - P_2\| = 1$$

Teorema.

Sean S_1 y S_2 subespacios de $\mathbb{R}^n$,

si $\dim(S_1) \neq \dim(S_2)$, entonces existe un vector en el subespacio de dimensión mayor, que es ortogonal al subespacio de dimensión menor y por consiguiente

$$\|P_1 - P_2\| = 1$$

Demostración.

La idea es hacer una demostración por contradicción, al suponer que si

$$\dim(S_1) > \dim(S_2)$$

tendremos que

$$S_2^{\perp} \cap S_1 \neq \{0\}$$

Supondremos entonces que

$$S_2^{\perp} \cap S_1 = \{0\}$$

entonces como

$$\dim(S_2^{\perp} + S_1) = \dim(S_2^{\perp}) + \dim(S_1)$$

tenemos

$$\dim(S_2^{\perp} + S_1) \leq n$$

y al sustituir obtenemos

$$[n - \dim(S_2)] + \dim(S_1) \leq n$$

de donde

$$\dim(S_1) \leq \dim(S_2)$$

lo que contradice a la hipótesis.

Por lo que

$$S_2^{\perp} \cap S_1 \neq \{0\}$$

de donde existe

$$x \neq 0, \text{ tal que } x \in S_2^{\perp} \cap S_1$$

y

$$\begin{aligned} (P_1 - P_2)x &= P_1x - P_2x \\ &= x - 0 \\ &= x \end{aligned}$$

de donde tenemos que

$$\|P_1 - P_2\| = 1.$$

Observemos también que si

$$S_2^{\perp} \cap S_1 \neq \{0\}$$

entonces, al tomar

$$x \in S_2^{\perp} \cap S_1; \quad x \neq 0$$

tenemos que

$$P_2^{\perp} P_1 x = x$$

entonces

$$\| P_2^{\perp} P_1 \| \geqslant 1$$

pero como

$$\| P_2^{\perp} P_1 z \| \leqslant \| P_2^{\perp} \| \cdot \| P_1 z \| \leqslant \| P_1 z \| \leqslant \| z \|$$

entonces

$$\| P_2^{\perp} P_1 \| \leqslant 1$$

de donde

$$\| P_2^{\perp} P_1 \| = 1$$

El resultado siguiente garantiza que la implicación en sentido opuesto es válido también

<u>Teorema.</u> $\| P_2^{\perp} P_1 \| = 1$ si y sólo si $S_2^{\perp} \cap S_1 \neq \{0\}$

<u>Demostración.</u>

Supongamos que $\| P_2^{\perp} P_1 \| = 1$, entonces existe $x \neq 0$ tal que

$$P_2^{\perp} P_1 x = x$$

Si $y = P_1 x$, tenemos que

$$P_2^{\perp} y = x$$

si $y \notin S_2^{\perp}$, entonces

$$\|x\| < \|y\|,$$

es decir,

$$\|P_1 x\| > \|x\|$$

lo cual no puede ocurrir ya que

$$\|P_1 x\| \leq \|x\|$$

siempre.

Entonces $$y \in S_2^{\perp}$$

y $$P_2^{\perp} y = y = x$$

así que $$x \in S_2^{\perp}$$

y entonces

$$P_1 x = x$$

de donde

$$x \in S_1 \Rightarrow S_2^{\perp} \cap S_1 \neq \{0\}$$

El caso cuando

$$P_2^{\perp} P_1 x = -x$$

no se puede dar, pues se llega a que

$$P_1 x = -x$$

de donde

$$P_1 x = x = -x$$

y se obtiene que

$$x = 0$$

La implicación en el otro sentido ya se demostró en el resultado anterior. ∎

Lo desarrollado hasta este momento lo resume el siguiente:

Teorema.

Si $$\|P_1 - P_2\| = 1$$

entonces $$\|P_2^{\perp} P_1\| = 1 \quad \text{o} \quad \|P_1^{\perp} P_2\| = 1$$

y además

$$\|P_2^{\perp} P_1\| < 1 \text{ si y sólo si } S_2^{\perp} \cap S_1 = \{0\}$$

si $$\|P_1^{\perp} P_2\| < 1 \text{ si y sólo si } S_1^{\perp} \cap S_2 = \{0\}$$

Los resultados anteriores pueden interpretarse desde otro punto de vista al introducir como hipótesis la dimensión de los subespacios.

CASO AGUDO: $\|P_1 - P_2\| < 1$.

Teorema.

Si $$\dim(S_1) = \dim(S_2),$$

entonces

$$S_1^{\perp} \cap S_2 \neq \{0\} \text{ si y sólo si}$$
$$S_2^{\perp} \cap S_1 \neq \{0\}.$$

Demostración.

Supongamos que
$$S_1^{\perp} \cap S_2 \neq \{0\}$$

sabemos, por otro lado, que
$$\dim(S_2 + S_1^{\perp}) + \dim(S_2 \cap S_1^{\perp}) =$$
$$= \dim(S_2) + \dim(S_1^{\perp})$$

entonces
$$\dim(S_2 + S_1^{\perp}) + \dim(S_2 \cap S_1^{\perp}) =$$
$$= \dim(S_2) + [n - \dim(S_1)]$$

y
$$\dim(S_2 + S_1^{\perp}) + \dim(S_2 \cap S_1^{\perp}) = n$$

de donde
$$\dim(S_2 + S_1^{\perp}) = n - \dim(S_2 \cap S_1^{\perp})$$
$$< n$$

entonces
$$(S_2 + S_1^{\perp})^{\perp} \neq \{0\}$$

Además
$$(S_2 + S_1^{\perp})^{\perp} \subset S_2^{\perp} \cap S_1$$

porque
$$S_2 \subset S_2 + S_1^{\perp}$$
$$S_1^{\perp} \subset S_2 + S_1^{\perp}$$

entonces
$$x \in (S_2 + S_1^{\perp})^{\perp}$$
$$\Rightarrow \quad x^t z = 0$$

para toda $z \in (S_2 + S_1^{\perp})$

para toda $z \in S_2$ tenemos que

$$x^t z = 0; \quad x \in S_2^{\perp}$$

para toda $z \in S_1^{\perp}$ tenemos que

$$x^t z = 0; \quad x \in (S_1^{\perp})^{\perp} = S_1$$

de donde

$$x \in S_2^{\perp} \cap S_1$$

es decir

$$(S_2^{\perp} \cap S_1) \neq \{0\}$$

La afirmación en el otro sentido vale por simetría. ∎

El resultado siguiente contiene al anterior:

Teorema.

Si

$$\dim(S_1) = \dim(S_2)$$

entonces

$$\dim(S_2^{\perp} \cap S_1) = \dim(S_1^{\perp} \cap S_2)$$

Demostración.

Supongamos que

$$\dim(S_2^{\perp} \cap S_1) = k \qquad y$$

$$\dim(S_1) = k + 1$$

Si $k = 0$, no hay nada que demostrar pues la afirmación se sigue del resultado anterior.

Así, supongamos que $k \neq 0$, y sean

$$\{b_1, b_2, \ldots, b_k\} \text{ una base}$$

para $S_2^{\perp} \cap S_1$,

y si
$$B = [\, b_1 | b_2 | \cdots | b_k]$$

entonces
$$S_2^{\perp} \cap S_1 = \mathcal{R}(B).$$

Sean
$$\{w_1, w_2, \ldots, w_l\}$$

tales que
$$\{b_1, b_2, \ldots, b_k, w_1, w_2, \ldots, w_l\}$$

es una base para S_1, es decir,
$$S_1 = \mathcal{R}([\, B | W]\,) = \mathcal{R}(B) + \mathcal{R}(W)$$

donde
$$\mathcal{R}(B) \cap \mathcal{R}(W) = \{0\}$$

con
$$W = [\, w_1 | w_2 | \ldots | w_l]$$

Si
$$\{z_1, z_2, \ldots, z_{k+1}\}$$

es una base de S_2, es decir, $S_2 = \mathcal{R}(Z)$

con
$$Z = [\, z_1 | z_2 | \ldots | z_{k+l}]$$

entonces como
$$\mathcal{R}(B) = S_2^{\perp} \cap S_1$$

tenemos
$$B^t Z \equiv 0$$

y como
$$S_1^{\perp} = N \begin{bmatrix} W^t \\ \cdots\cdots \\ B^t \end{bmatrix}$$

resulta que

$$S_2 \cap S_1^{\perp} = \mathcal{R}(Z) \cap N \begin{bmatrix} W^t \\ \cdots\cdots \\ B^t \end{bmatrix}$$

$$= \{y / y = Z\tilde{y} \text{ y } \begin{bmatrix} W^t \\ \cdots\cdots \\ B^t \end{bmatrix} y = 0\}$$

$$= \{\tilde{y} / \begin{bmatrix} W^t \\ \cdots\cdots \\ B^t \end{bmatrix} Z\tilde{y} = 0\}$$

$$= \{\tilde{y} / \begin{bmatrix} W^t Z \\ \cdots\cdots \\ B^t Z \end{bmatrix} \tilde{y} = 0\}$$

como

$$B^t Z = 0$$

$$S_2 \cap S_1^{\perp} = \{\tilde{y} / \begin{bmatrix} W^t Z \\ \cdots\cdots \\ 0 \end{bmatrix} \tilde{y} = \begin{bmatrix} 0 \\ \cdots\cdots \\ 0 \end{bmatrix} \}$$

$$= \{\tilde{y} / W^t Z \tilde{y} = 0\}$$

$$= \{\tilde{y} / A\tilde{y} = 0;\ A = W^t Z\}$$

de donde

$$S_2 \cap S_1^{\perp} = N(A) = N(W^t Z)$$

como Z es de rango máximo, y rango $(W^tZ) = $ rango $(W^t) = 1$, y como W tiene columnas independientes entonces W^tZ tiene renglones independientes y así

$$\text{rango } (W^tZ) = 1 = \text{rango } (W^t)$$

Ahora, como

$$\text{rango}(W^tZ) + \dim(N(W^tZ)) = k + 1$$

tenemos $$1 + \dim(S_2 \cap S_1^{\perp}) = k + 1$$

por lo tanto
$$\dim(S_1^{\perp} \cap S_2) = k. \qquad \blacksquare$$

Los resultados anteriores se pueden resumir con el siguiente:

<u>Teorema.</u>

Si $$\dim(S_1) = \dim(S_2)$$

entonces las siguientes afirmaciones son equivalentes

i) $\| P_2^{\perp} P_1 \| < 1$

ii) $\| P_1^{\perp} P_2 \| < 1$

iii) $\| P_1 - P_2 \| < 1$

y cuya validez se infiere directamente de resultados anteriores y del hecho que

$$\begin{aligned} P_2^{\perp} P_1 &= (I - P_2) P_1 \\ &= P_1 - P_2 P_1 \\ &= P_1^2 - P_2 P_1 \\ &= (P_1 - P_2) P_1 \end{aligned}$$

Al tomar normas se obtiene

$$\begin{aligned} \| P_2^{\perp} P_1 \| &\leq \| P_1 - P_2 \| \cdot \| P_1 \| \\ &\leq \| P_1 - P_2 \| . \qquad \blacksquare \end{aligned}$$

A continuación, veremos con más detalle el significado de

$$\| P_1 - P_2 \| < 1.$$

Para ello notemos que

Lema.

Si

$$\| P_2^{\perp} P_1 \| < 1$$

entonces

$$\dim(P_2(S_1)) = \dim(S_1)$$

y por consiguiente

$$\dim(S_1) < \dim(S_2).$$

Demostración.

Como

$$S_2^{\perp} \cap S_1 = \{0\}$$

entonces

para toda $x \in S_1$

$$x \neq 0, \quad P_2 x \neq 0 \;\; \Rightarrow P_2,$$

restringida a S_1 es uno a uno, y por consiguiente

$$\dim(P_2(S_1)) = \dim(S_1)$$

pero como

$$P_2(S_1) \subset S_2$$

entonces

$$\dim(S_1) < \dim(S_2).$$

∎

Al tomar este resultado como base, podemos concluir con lo siguiente:

<u>Teorema.</u>

Si $$\|P_1 - P_2\| < 1$$

entonces

i) $\|P_2^{\perp} P_1\| < \|P_1 - P_2\| < 1$

ii) $\|P_1^{\perp} P_2\| < \|P_1 - P_2\| < 1$

iii) $\dim(S_1) = \dim(S_2)$

El resultado anterior, sugiere la siguiente pregunta:

¿Bajo qué condiciones ocurre que

$$\|P_2^{\perp} P_1\| = \|P_1 - P_2\| < 1?$$

Para responder, observemos que

$$\begin{aligned}
\|(P_1-P_2)w\|^2 &= \|P_2(P_1-P_2)w + P_2^{\perp}(P_1-P_2)w\|^2 \\
&= \|P_2(P_1-P_2)w\|^2 + \|P_2^{\perp}(P_1-P_2)w\|^2 \\
&= \|-P_2(I-P_1)w\|^2 + \|P_2^{\perp}P_1 w\|^2 \\
&< \|P_2P_1^{\perp}\|^2 \cdot \|P_1^{\perp}w\|^2 + \|P_2^{\perp}P_1\|^2 \cdot \|P_1 w\|^2
\end{aligned}$$

y como

$$\|P_2P_1^{\perp}\| = \|P_1^{\perp}P_2\|$$

tenemos

$$\|(P_1-P_2)w\|^2 \leq \|P_1^{\perp}P_2\|^2 \cdot \|P_1^{\perp}w\|^2 + \|P_2^{\perp}P_1\|^2 \cdot \|P_1 w\|^2$$

$$\leq \max\{\|P_1^{\perp}P_2\|^2,\ \|P_2^{\perp}P_1\|^2\}(\|P_1^{\perp}w\|^2 +$$

$$+ \|P_1 w\|^2)$$

$$\leq \max\{\|P_1^{\perp}P_2\|^2,\ \|P_2^{\perp}P_1\|^2\}\|w\|^2$$

de donde

$$\|P_1-P_2\| \leq \max\{\|P_1^{\perp}P_2\|,\ \|P_2^{\perp}P_1\|\}$$

Con lo anterior se puede enunciar el siguiente:

<u>**Teorema.**</u>

$$\|P_1-P_2\| \leq \max\{\|P_2^{\perp}P_1\|,\ \|P_1^{\perp}P_2\|\}$$

Con los dos últimos resultados obtenemos:

<u>**Teorema.**</u>

Si

$$\|P_1-P_2\| < 1$$

entonces

$$\|P_2^{\perp}P_1\| = \|P_1^{\perp}P_2\| = \|P_1 - P_2\|.$$

Demostración.

Supongamos que

$$\max\{\|P_2^{\perp}P_1\|,\ \|P_1^{\perp}P_2\|\} = \|P_2^{\perp}P_1\|,$$

entonces, por los resultados anteriores obtenemos que

$$\|P_1^{\perp}P_2\| < \|P_1-P_2\| < \|P_2^{\perp}P_1\|$$

pero, al cambiar los papeles de S_1 y S_2, obtenemos

$$\|P_2^{\perp}P_1\| < \|P_1-P_2\| < \|P_1^{\perp}P_2\|$$

de donde obtenemos el resultado deseado. ∎

Nota.

Los desarrollos anteriores nos permitirán pasar al siguiente problema que consiste en obtener las cotas de perturbación.

OBTENCION DE LAS COTAS DE PERTURBACION.

Para la obtención de las cotas de perturbación es necesario considerar la forma en que las proyecciones vienen dadas, ya que los subespacios S_1 y S_2 son por lo general los Núcleos o Rangos de matrices.

Para comenzar, consideremos el caso en que S_1 y S_2 tienen bases ortonormales.

Sean

$$\{u_1, u_2, \ldots, u_k\}$$

y

$$\{v_1, v_2, \ldots, v_k\}$$

bases ortonormales de S_1 y S_2 respectivamente, entonces

$$P_1 = u_1 u_1^t + u_2 u_2^t + \ldots + u_k u_k^t = U_1 U_1^t$$

$$P_2 = v_1 v_1^t + v_2 v_2^t + \ldots + v_k v_k^t = V_1 V_1^t$$

donde

$$U_1 = [\, u_1 | u_2 | \ldots | u_k] \text{ es de } n \times k$$

$$V_1 = [\, v_1 | v_2 | \ldots | v_k] \text{ es de } n \times k$$

y U_1 y V_1 son tales que

$$U_1^t \, U_1 = I_k ; \quad V_1^t \, V_1 = I_k$$

y tenemos que

$$\| P_1 - P_2 \| = \| U_1 U_1^t - V_1 V_1^t \|$$

$$\| P_1^{\perp} \, P_2 \| = \| (I - U_1 U_1^t) V_1 V_1^t \|$$

$$\| P_2^{\perp} \, P_1 \| = \| (I - V_1 V_1^t) U_1 U_1^t \|$$

y como

$$\dim(S_1) = \dim(S_2)$$

las tres cantidades son iguales y se calcula aquélla cuya evaluación sea la más fácil.

Las expresiones de las tres normas anteriores pueden simplificarse, debido a que

$$\|U_1 x\| = \|x\|$$

$$\|V_1 y\| = \|y\|$$

para cualesquiera x e y en $\mathbb{R}^n$.

Sin embargo, es más conveniente proceder como se indica a continuación:

Sean

$$\{u_{k+1}, u_{k+2}, \ldots, u_n\}$$

tales que

$$\{u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_n\}$$

es una base ortonormal para $\mathbb{R}^n$ y análogamente con

$$\{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\} .$$

Si tomamos

$$U = [\underset{k}{U_1} | \underset{n-k}{U_2}]$$

$$= [u_1 | u_2 | \ldots | u_k | u_{k+1} | \ldots | u_n]$$

entonces

$$P_1 = U_1 U_1^t = [U_1 | 0] \left[\begin{array}{c} U_1^t \\ \hline 0 \end{array} \right]$$

$$= [\,U_1 \,|\, U_2\,] \begin{bmatrix} I_k \\ \cdots \\ 0 \end{bmatrix} [\,I_k \,|\, 0\,] \begin{bmatrix} U_1^t \\ \cdots \\ U_2^t \end{bmatrix}$$

$$= U \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] U^t$$

donde $U_{n\times n}$ es una matriz ortogonal y análogamente

$$P_2 = V \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] V^t$$

con $V_{n\times n}$, una matriz ortogonal.

Con estas dos expresiones para P_1 y P_2 se calcula

$$\| P_1 - P_2 \| .$$

Entonces

$$P_1 - P_2 = U \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] U^t - V \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] V^t$$

y al aplicar normas

$$\| P_1 - P_2 \| = \left\| \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] U^t - U^t V \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array}\right] V^t \right\|$$

$$\|P_1-P_2\| = \left\| \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} U^tV - U^tV \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \right\|$$

$$= \left\| \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} Q - Q \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \right\|$$

$$= \left\| \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \begin{bmatrix} Q_{11} & \vdots & Q_{12} \\ \cdots & & \cdots \\ Q_{21} & \vdots & Q_{22} \end{bmatrix} - \begin{bmatrix} Q_{11} & \vdots & Q_{12} \\ \cdots & & \cdots \\ Q_{21} & \vdots & Q_{22} \end{bmatrix} \begin{bmatrix} I_k & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} \right\|$$

$$= \left\| \begin{bmatrix} Q_{11} & \vdots & Q_{12} \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{bmatrix} - \begin{bmatrix} Q_{11} & \vdots & 0 \\ \cdots & & \cdots \\ Q_{21} & \vdots & 0 \end{bmatrix} \right\|$$

$$= \left\| \begin{bmatrix} 0 & \vdots & Q_{12} \\ \cdots & & \cdots \\ -Q_{21} & \vdots & 0 \end{bmatrix} \right\|$$

con

$$Q = U^tV = \begin{bmatrix} U_1^t \\ \cdots \\ U_2^t \end{bmatrix} [V_1 | V_2]$$

$$= \begin{bmatrix} U_1V_1 & \vdots & U_1^tV_2 \\ \cdots & & \cdots \\ U_2^tV_1 & \vdots & U_2^tV_2 \end{bmatrix}$$

de donde

$$Q_{12} = U_1^t\, V_2$$

$$Q_{21} = U_2^t\, V_1$$

y así

$$\|P_1 - P_2\| = \left\| \left[\begin{array}{c:c} 0 & U_1^t V_2 \\ \hdashline -U_2^t V_1 & 0 \end{array} \right] \right\|$$

A continuación, calcularemos una expresión para

$$\|P_1^{\perp} P_2\|$$

Así pues, como

$$P_1^{\perp} = U \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & I_{n-k} \end{array} \right] U^t$$

tenemos

$$P_1^{\perp} P_2 = U \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & I_{n-k} \end{array} \right] U^t V \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array} \right] V^t$$

y, al aplicar normas

$$\|P_1^{\perp} P_2\| = \left\| \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & I_{n-k} \end{array} \right] Q \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array} \right] \right\|$$

$$= \left\| \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & I_{n-k} \end{array} \right] \left[\begin{array}{c:c} Q_{11} & Q_{12} \\ \hdashline Q_{21} & Q_{22} \end{array} \right] \left[\begin{array}{c:c} I_k & 0 \\ \hdashline 0 & 0 \end{array} \right] \right\|$$

$$= \left\| \left[\begin{array}{c:c} 0 & 0 \\ \hdashline Q_{21} & 0 \end{array} \right] \right\|$$

de donde $$\|P_1^{\perp} P_2\| = \|Q_{21}\|$$

y análogamente tenemos para $\|P_2^{\perp} P_1\|$ que

$$\|P_2^{\perp} P_1\| = \|Q_{12}\|$$

Estos resultados nos indican que para calcular

$$\|P_1 - P_2\|$$

basta calcular

$$\|Q_{12}\| = \|U_1^t V_2\|$$

o

$$\|Q_{21}\| = \|U_2^t V_1\|$$

Además, como sólo estamos considerando el caso en que

$$\dim(S_1) = \dim(S_2)$$

entonces

$$\|P_1 - P_2\| = \|P_1^{\perp} P_2\| = \|P_2^{\perp} P_1\| .$$

Podemos concluir con el siguiente resultado para matrices ortogonales:

<u>Teorema.</u>

Si $Q_{n\times n}$ es una matriz ortogonal y es tal que

$$Q = \begin{array}{c c} \left[\begin{array}{c:c} Q_{11} & Q_{12} \\ \hdashline Q_{21} & Q_{22} \end{array}\right] & \begin{array}{l} k \\ n-k \end{array} \\ \begin{array}{cc} k & \quad n-k \end{array} & \end{array}$$

entonces

$$\|Q_{21}\| = \|Q_{12}\| .$$

Este resultado también se obtiene al tomar S_1 como el subespacio generado por las k primeras columnas de Q y a S_2 como el subespacio generado por los k primeros vectores de la base canónica de $\mathbb{R}^n$.

De esta forma se puede ver que

$$\|P_2^{\perp} P_1\| = \|Q_{21}\|$$

y
$$\|P_1^{\perp} P_2\| = \|Q_{12}\|$$

Pues si

$$x = Qa = \left[\begin{array}{c:c} Q_{11} & Q_{12} \\ \hdashline Q_{21} & Q_{22} \end{array}\right] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

entonces, por ser Q ortogonal, tenemos que

$$P_1 x = \left[\begin{array}{c} Q_{11}\, a_1 \\ \hdashline Q_{21}\, a_1 \end{array}\right]$$

de donde

$$P_2^{\perp}(P_1x) = \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & I_{n-k} \end{bmatrix} \begin{bmatrix} Q_{11} & a_1 \\ \cdots & \cdots \\ Q_{21} & a_1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \cdots \cdots \\ Q_{21} \; a_1 \end{bmatrix} = \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ Q_{21} & \vdots & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

es decir

$$P_2^{\perp} \; P_1 = \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ Q_{21} & \vdots & 0 \end{bmatrix}$$

y al aplicar normas

$$\| P_2^{\perp} \; P_1 \| = \| Q_{21} \|$$

El caso cuando

$$\| P_1^{\perp} \; P_2 \| = \| Q_{12} \|$$

es análogo al anterior.

Trataremos ahora el caso cuando los subespacios S_1 y S_2 son los Rangos de las matrices A y B:

Así pues, sean

$$S_1 = \mathcal{R}(A) \quad \text{y} \quad S_2 = \mathcal{R}(B)$$

donde

$$B = A + E$$

y $$\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(B))$$

entonces $$P_1 = AA^+ \text{ y } P_2 = BB^+$$

Necesitamos ahora obtener una cota que "vaya" a cero y que sea lo más simple posible. Podemos ver, que si $\|E\|$ es pequeña los subespacios están cercanos y la diferencia

$$\|P_1 - P_2\|$$

tiende a cero:

$$\|P_1 - P_2\| \to 0$$

Pero aquí surge una dificultad, pues como no se sabe si A y B tienen columnas ortogonales, el cálculo de

$$\|P_1 - P_2\|$$

no es fácil. De aquí que partiendo del hecho que

$$\|P_1 - P_2\| = \|P_2^{\perp} P_1\| = \|P_1^{\perp} P_2\|$$

probaremos con evaluar

$$\|P_2^{\perp} P_1\| \quad \text{y} \quad \|P_1^{\perp} P_2\|$$

para encontrar cuál es la expresión más simple que nos convenga.

Así pues, al desarrollar, tenemos

$$P_2^{\perp} P_1 = P_2^{\perp} AA^+ = P_2^{\perp}(B - E)\, A^+$$

$$= P_2^{\perp} BA^+ - P_2^{\perp} EA^+$$

Como $$P_2^{\perp} B = 0$$

tenemos $$P_2^{\perp} P_1 = - P_2^{\perp} E A^+$$

y $$\| P_2^{\perp} P_1 \| \leq \| E A^+ \| \leq \| E \| \cdot \| A^+ \|$$

También
$$P_1^{\perp} P_2 = P_1^{\perp} B B^+ = P_1^{\perp} (A + E) B^+$$
$$= P_1^{\perp} A B^+ + P_1^{\perp} E B^+$$

como $$P_1^{\perp} A = 0$$

tenemos
$$P_1^{\perp} P_2 = P_1^{\perp} E B^+$$

y $$\| P_1^{\perp} P_2 \| \leq \| E B^+ \| \leq \| E \| \cdot \| B^+ \|$$

Así observamos que es más fácil obtener

$$\| P_1 - P_2 \| = \| P_2^{\perp} P_1 \| = \| P_1^{\perp} P_2 \| < 1$$

al pedir que
$$\| E \| \cdot \| A^+ \| < 1$$

Obsérvese también que vuelve a aparecer esta condición que ya se había presentado en términos de valores singulares. Y esta vez, la interpretación que se le da es la siguiente.

No deben existir vectores en $\mathcal{R}(A)$ que son ortogonales a $\mathcal{R}(B)$ y viceversa, pues de otra forma los subespacios no estarían cercanos, y además las matrices A y B deben de diferir poco.

Consideremos el caso, ahora, cuando los subespacios S_1 y S_2 son los espacios ortogonales a los Núcleos de las matrices A y B:

Así, pues, sean

$$S_1 = N(A)^{\perp} \quad y \quad S_2 = N(B)^{\perp}$$

donde

$$B = A + E$$

y

$$\dim(R(A)) = \dim(R(B))$$

entonces,

$$P_1 = A^{+} A \quad y \quad P_2 = B^{+} B$$

y al desarrollar, obtenemos

$$P_2^{\perp} P_1 = P_2^{\perp} A^{+} A = P_2^{\perp} A^{t}(A^{t})^{+}$$

$$= P_2^{\perp} (B^{t} - E^{t})(A^{t})^{+}$$

$$= P_2^{\perp} B^{t}(A^{t})^{+} - P_2^{\perp} E^{t}(A^{t})^{+}$$

pero

$$P_2^{\perp} B^{t} = 0$$

entonces

$$P_2^{\perp} P_1 = - P_2^{\perp} E^{t}(A^{t})^{+}$$

Al aplicar normas

$$\|P_2^{\perp} P_1\| < \| \quad \| \cdot \|A^{+}\|$$

Análogamente

$$P_1^{\perp} P_2 = P_1^{\perp} B^{+} B = P_1^{\perp} B^{t} (B^{t})^{+}$$

$$= P_1^{\perp} (A + E)^{t} (B^{t})^{+}$$

$$= P_1^{\perp} E^{t} (B^{t})^{+}$$

ya que $\quad P_1^{\perp} A = 0$

Al aplicar normas

$$\|P_1^{\perp} P_2\| < \|E\| \quad \|B^{+}\|$$

Con estos resultados podemos concluir que la condición para que los Rangos de A y B estén cercanos es la misma que se necesita para que sus Núcleos lo estén:

<u>COTAS DE PERTURBACION DE LA SOLUCION.</u>

Para poder evaluar los efectos que sufre la solución del problema de Cuadrados Mínimos cuando se ve afectada por perturbaciones, utilizaremos, como ya se indicó el enfoque geométrico que acabamos de desarrollar.

Tenemos el ya conocido problema, en donde

$$x = A^{+} b$$

$$y = B^{+} b$$

con

$$A + E = B$$

$$\text{rango } (A) = \text{rango } (B)$$

y

$$\|E\| \cdot \|A^{+}\| < 1$$

Para conocer los efectos de las perturbaciones en la solución, se requiere calcular

$$\|x - y\|$$

La idea en la que nos basaremos para hacerlo es la siguiente:

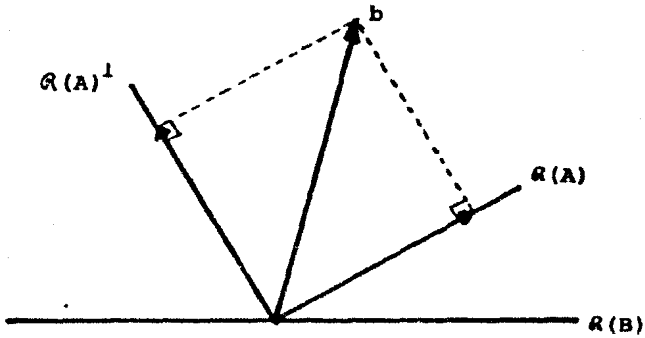
$$b = P_{\mathcal{R}(A)}\, b + P_{\mathcal{R}(A)^{\perp}}\, b$$

entonces

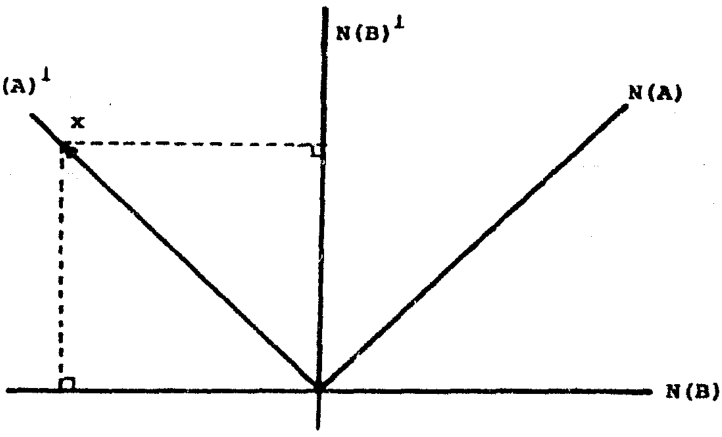
$$y = B^{+}b = B^{+}(P_{\mathcal{R}(A)}b + P_{\mathcal{R}(A)^{\perp}}\, b)$$

$$y = B^{+}\, P_{\mathcal{R}(A)}b + B^{+}\, P_{\mathcal{R}(A)^{\perp}}\, b$$

y gráficamente tenemos



Además



$$x = A^{+}b = P_{N(B)}x + P_{N(B)^{\perp}}x$$

$$= P_{N(B)}A^{+}b + P_{N(B)^{\perp}}A^{+}b$$

Y entonces, al sustituir en (y - x), obtenemos

$$y - x = B^{+} P_{\mathcal{R}(A)} b + B^{+} P_{\mathcal{R}(A)^{\perp}} b -$$

$$P_{N(B)} A^{+} b - P_{N(B)^{\perp}} A^{+} b$$

como

$$P_{N(B)^{\perp}} = B^{+} B$$

se pueden asociar los términos

$$B^{+} P_{\mathcal{R}(A)} \quad y \quad P_{N(B)^{\perp}} A^{+}$$

y obtener

$$(y - x) = B^{+} P_{\mathcal{R}(A)} b - P_{N(B)^{\perp}} A^{+} b +$$

$$B^{+} P_{\mathcal{R}(A)^{\perp}} b - P_{N(B)^{\perp}} A^{+} b$$

$$= B^{+} (P_{\mathcal{R}(A)} - BA^{+}) b +$$

$$B^{+} P_{\mathcal{R}(A)^{\perp}} b - P_{N(B)} A^{+} b$$

y considerando que

$$P_{\mathcal{R}(A)} = A A^{+}$$

$$B^{+} P_{\mathcal{R}(B)} = B^{+}$$

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

se llega a lo siguiente:

$$(y - x) = B^{+}(A A^{+} - B A^{+})\, b +$$

$$B^{+} P_{R(B)} P_{R(A)^{\perp}}\, b -$$

$$P_{N(B)} P_{N(A)^{\perp}} A^{+} b$$

de donde

$$(y - x) = B^{+} (A - B) A^{+} b +$$

$$B^{+} P_{R(B)} P_{R(A)^{\perp}}\, b -$$

$$P_{N(B)} P_{N(A)^{\perp}} A^{+} b$$

Esta última expresión puede evaluarse fácilmente cuando se toman normas.

Lo anterior nos conduce a enunciar el siguiente resultado, fundamental en este trabajo.

Teorema.

Sean A y B dos matrices de $m \times n$. Si satisfacen las hipótesis siguientes:

i) $B = A + E$

ii) rango (B) = rango (A)

iii) $\|E\| \cdot \|A^{+}\| < 1$

entonces

$$\|B^{+} - A^{+}\| \leq 2\, \frac{\|E\| \cdot \|A^{+}\|^{2}}{1 - \|E\| \cdot \|A^{+}\|} + \|E\| \cdot \|A^{+}\|^{2}$$

Si $\qquad y = B^{+} b \quad y \quad x = A^{+} b$

entonces

$$\|y - x\| < \frac{\|E\| \cdot \|A^{+}\|}{1-\|E\| \cdot \|A^{+}\|} \cdot \|x\| \; +$$

$$\frac{\|E\| \cdot \|A^{+}\|^{2}}{1-\|E\| \cdot \|A^{+}\|} \cdot \|r\| + \|E\| \cdot \|A^{+}\| \cdot \|x\|$$

Demostración.

Es inmediata a partir de resultados anteriores y de que

$$\|B^{+}\| < \frac{\|A^{+}\|}{1-\|E\| \; \|A^{+}\|}$$

como ya se demostró anteriormente. ∎

El resultado anterior, en términos de la condición de A, se convierte en:

Como $\qquad k(A) = \|A\| \cdot \|A^{+}\|$

si llamamos

$$\tilde{k}(A) = \frac{k(A)}{1-\|E\| \cdot \|A^{+}\|}$$

y si además

$$\alpha(A) = \frac{\|E\|}{\|A\|}$$

$$\beta(b) = \frac{\|e\|}{\|b\|}$$

$$\gamma(b) = \frac{\|b\|}{\|A\| \cdot \|x\|} \leq \frac{\|b\|}{\|Ax\|}$$

$$\rho(b) = \frac{\|r\|}{\|A\| \cdot \|x\|} \leq \frac{\|r\|}{\|Ax\|}$$

tenemos:

<u>Teorema.</u>

Con las hipótesis del resultado anterior y si

$$z = B^{+}(b+e)$$

$$x = A^{+} b \; ; \quad x \neq 0$$

entonces

$$\frac{\|z-x\|}{\|x\|} \leq \tilde{k}(A) \left[\alpha(A) + k(A)\rho(b)\alpha(A) + \gamma(b)\beta(b) \right]$$

$$+ k(A)\alpha(A)$$

y como

$$\tilde{k}(A) \geq k(A)$$

tenemos

$$\frac{\|z-x\|}{\|x\|} \leq \tilde{k}(A) \left[(2 + k(A)\rho(b))\alpha(A) + \gamma(b)\beta(b) \right]$$

Nota.

Obsérvese que las cotas obtenidas coinciden con las del capítulo anterior, con la diferencia de que aparece $\tilde{k}(A)$ en lugar de $k(A)$.

BIBLIOGRAFIA.

[1] Björck, Åke y Golub, G.H. "Numerical Methods for Computing Angles between Linear Subspaces". *Mathematics of Computation* 27, (1973) 579-594.

[2] Davis, Chandler y Kahan, W.M. "The Rotation of Eigenvectors by a Perturbation, III" *S.I.A.M. J. Numerical Analysis 7 No. 1* 1-46.

[3] Faddeev, D.K. y Faddeeva, V.N. Computational Methods of Linear Algebra San Francisco, W.H. Freeman and Company, 1963, 621 p.p.

[4] Kato, Tosio Perturbation Theory for Linear Operators. New York, Springer-Verlag, 1966.

[5] Lang, Serge Linear Algebra 2a. edición Filipinas, Addisoon-Wesley, 1971, 400 p.p.

[6] Lawson, Charles L, y Hanson Richard J. Solving Least Squares Problems. New Jersey, Prentice-Hall, 1974, 340 p.p.

[7] Máltsev A.I. Fundamentos de Algebra Lineal. Traducido del ruso por Carlos Vega.Moscú, MIR, 1972 400 p.p.

[8] Rice, John R. Matrix Computations and Mathematical Software. E.U.A. McGraw-Hill, 1981 248 p.p.

[9] Stewart, G.W. Introduction to Matrix Computations. New York, Academic-Press 1973, 441 p.p.

[10] Stewart G.W. "On the Perturbation of Pseudo-inverses, projections and linear least squeares prcblems". *S.I.A.M. Review Vol. 19 No. 4* Octubre 1977 634-662

[11] Strang, G. Linear Algebra and its Applications. New York, Academic Press, 1976 374 p.p.